2014

People are Neither Compatibilists Nor Incompatibilists: They Maintain Distinct, Inconsistent Intuitions Regarding Determinism and Moral Responsibility

Stephanie C. Cerce
Lehigh University

People are Neither Compatibilists Nor Incompatibilists: They Maintain Distinct, Inconsistent Intuitions Regarding Determinism and Moral Responsibility

by

Stephanie C. Cerce

A Thesis

Presented to the Graduate and Research Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Psychology

Lehigh University

May 2014

Thesis is accepted and approved in partial fulfillment of the requirements for the Master of Science in the Psychology Department.

People are Neither Compatibilists Nor Incompatibilists: They Maintain Distinct, Inconsistent Intuitions Regarding Determinism and Moral Responsibility

Stephanie C. Cerce

_____
Date Approved

_____
Michael J. Gill, Ph.D.
Adviser

_____
Deborah J. Laible, Ph.D.
Committee Member

_____
Dominic J. Packer, Ph.D.
Committee Member

_____
Gordon B. Moskowitz, Ph.D.
Department Chair Person

Table of Contents

List of Figures

Abstract

The implications a deterministic world might have for moral responsibility has long been a topic of philosophical debate. While much current research asserts that people are either incompatibilists or compatibilists, we argue that people do not consistently apply these beliefs. Instead, we argue that people maintain both the intuition that determinism diminishes moral responsibility and the intuition that people can be held morally responsible even in the presence of deterministic forces. We presented three studies to test the hypothesis that people will ignore determinism information when 1) selectively reminded that the actor maintains free will or 2) motivated to punish. We found some evidence to support our claim: People lay blame despite the existence of deterministic factors. However, this effect depends on the nature of the scenario and preexisting characteristics of the individual. Future work is necessary to determine if there is an underlying consistency to this effect.

Philosophers have long been concerned with the implications a deterministic world has for moral responsibility (Aristotle, 350BCE; Hume, 1738; Kant, 1781; Strawson, 1963; Watson, 1987). Much debate has focused on what the *true state of the world* is and whether, if we live in a deterministic universe, we ought to ascribe responsibility for moral and immoral actions. Regardless of any metaphysical or normative claims, however, it would appear that everyday people have quite malleable views regarding when an individual is morally responsible. For example, extant research on blame has provided significant evidence that *intentional* immoral actions elicit heightened ascriptions of moral responsibility as compared to unintentional actions (Alicke, 2000; Lagnado & Channon, 2008; Pizarro, Uhlmann, & Salovey, 2003; Weiner, 2006; Woolfolk, Doris, & Darley, 2006). But, on the other hand, Gill and Cerce (2013) suggest a revision to this framework by showing how blame can be mitigated even when a transgression is intentional. This is with information suggesting that the actor's negative intentions were *implanted*, or established in her mind as a result of unfortunate formative life experiences. Implantation information implies a deterministic world in which an individual is not the sole architect of her dispositions. Indeed, Gill and Cerce reported that implantation information attenuates blame by creating the impression that the transgressor lacked free will over becoming a person with negative, immoral intentions. Taken together, these works suggest that people generally ascribe blame as if we do not live in a deterministic world, but when determinism is made salient—in the form of information about an actor's unfortunate history—their blame ascriptions are significantly reduced.

The purpose of the research below is to explore the possibility that blaming practices are more complex than implied by either the literature on perceived intentionality or by Gill and Cerce's (2013) work on character implantation. We are interested in the possibility that observers are inconsistent in their judgments of an immoral actor with an unfortunate history: Sometimes they will discount blame based on the deterministic implications of the historical information, and sometimes they will ignore the deterministic implications of the historical information and blame as if (or almost as if) that information was never presented. As will be elaborated below, we argue this can happen based on both informational and motivational factors. For example, the observer may consider information that results in holding on to the notion of the transgressor's human capacity for free will—focusing on the fact that the transgressor can still *choose to be a better person moving forward*, despite her lacking free will over the *initial acquisition* of negative dispositions. In this way, people perceive a continuum of free will: At times a transgressor can be seen as acting fully of his own free will and at other times his free will appears to be diminished as a result of deterministic forces (though not eradicated). Because of this continuum, it appears as though people maintain inconsistent intuitions[1] regarding whether determinism precludes moral responsibility—when, instead, perceptions of free will are a matter of gradation, resulting in varying responsibility ascriptions based on this perception. Therefore, in one context, people believe that determinism can mitigate control over becoming a certain type of (immoral)

[1] We subscribe to the definition of an intuition provided by Haidt (2001): "[An] intuition occurs quickly, effortlessly, and automatically, such that the outcome but not the process is accessible to consciousness…" (p. 818). However, we endorse the use of "intuition" as a matter of convenience: We do not have any evidence nor are we entirely confident that the judgments discussed in this paper are the result of an intuitive or deliberative process.

person, while, in another, that the actor can be held morally responsible for his negative deeds even if his history determined his immoral character traits.

Blame Depends on Perceived Intentionality

Much of the existing blame literature focuses on perceived intentionality and its implications for heightening blame. In the prototypical blame scenario, the actor produces a negative event and, in response, the observer experiences "negative reactive attitudes" (Strawson, 1963) toward the actor. These reactive attitudes include anger, dislike, resentment, and indignation, which are collectively labeled "blame." Prior literature emphasizes that it is the observer's perceptions of the *mental states and mental activities* of the actor that determine how much blame will be felt (see Malle, Guglielmo, & Monroe, 2011, for a review). Specifically, research has shown that perceptions of the transgressor's *desire* to commit an immoral deed (Lagnado & Channon, 2008; Malle & Knobe, 1997), *identification* with the action (Pizarro et al., 2003; Woolfolk et al., 2006), and volitional *control* (Alicke, 2000; Weiner, 2006) increase blame ascriptions. Perceived intentionality, then, is based on the confluence of several factors.

For instance, Malle & Knobe (1997) suggest that the core features of perceived intentionality are perceptions of an agent's *desire* to harm and perceptions that the agent *believed* her action would produce that harm. Consistent with this conceptualization, Lagnado & Channon (2008) had participants read scenarios with information regarding whether the act was done intentionally or unintentionally (e.g., *Sue dislikes her colleague and puts a virus on his computer to damage it* versus *Sue accidently puts a virus on her colleague's computer*), and whether the actor believed the behavior would produce the desired outcome (e.g., *Sue thinks the virus is likely to cause serious damage* versus *Sue*

thinks the virus is unlikely to cause serious damage). Results indicated main effects of intentionality and foreseeability, suggesting that participants ascribe greater blame to actors who both act on a desire to harm and believe their behavior will fulfill that desire.

In addition to perceptions of desire and foreseeability, observers also consider whether the actor "whole-heartedly" embraces the behavior, or *identifies* with the action (Pizarro, et al., 2003; Woolfolk, et al., 2006). Indeed, observers consider the actor's second-order desires, or his "higher-order acceptance or rejection of a desire or impulse," when making moral judgments (Pizarro et al., 2003; p. 267). A transgressor who commits a negative deed (e.g., *Martha steals from the store*), and whose second-order desire aligns with that action (e.g., *Martha is pleased to have done it.*), is considered extremely blameworthy. In one study, Pizarro et al. (2003) had participants rate the blameworthiness of deliberate or impulsive actors who committed transgressions. Results indicated that participants discounted blame for *impulsive* negative acts relative to *deliberate* negative acts. This was mediated by participants' beliefs about the actor's second-order desire: Participants' believed that an actor did not have a congruent second-order desire when committing an *impulsive* act, but he did have such a second-order desire when he had committed a *deliberate* act—resulting in differential blame ascriptions.

Finally, blame ascriptions are influenced by the extent to which the actor is considered able to volitionally change her behavior (Alicke, 2000; Weiner, 1970, 2006). For example, when there are capacity (e.g., mental illness) constraints, the actor's volitional control is diminished and she is blamed less. In Weiner and Kukla's (1970) study, participants learned about students who failed an exam. One group failed due to

lack of effort, the other due to lack of ability. Failure due to lack of effort is seen as controllable (*the student could have done better if he had chosen to study more*) whereas failure due to lack of ability is not (*even if the student had tried harder, he does not have the ability to do better*). Observers recognize that the student who fails due to lack of effort made a volitional choice not to study and, thus, that student is blamed more. In contrast, the student with lack of ability was not able to choose to do otherwise, and is seen as less blameworthy.

These various perceptions of desire, identification, and control are all part of the umbrella concept of perceived intentionality. And, arguably, observers consider these mental states and mental activities because the transgressor's intentionality provides information regarding his or her moral character and thus the likelihood he or she will engage in negative behavior in the future (Pizarro & Tannenbaum, 2011).

Blame Implications of Perceived Intentionality Can Be Nullified by Determinism Information Regarding an Unfortunate History

Gill and Cerce (2013) argue that the well-documented link between perceived intentionality and blame can be broken with information regarding unfortunate formative forces in the actor's history. These forces suggest that the transgressor's negative intentions (*he wants to dominate and control others*) were implanted (*as a result of his painful life, full of humiliation*), and thus that the transgressor never "set out" to have those intentions. Instead, the intentions are a result of deterministic forces, undermining the perception that they are a manifestation of "free will," fully under the control of the actor. We tested these ideas in an experiment. One condition described an impulsive (but regretful) bully, whereas the other two conditions described an intentional bully whose

bad behavior possessed all the features of intentionality noted above: Desire, foreseeability, deliberateness, identification, and control. In one of the two intentional conditions, participants learned about the bully's unfortunate history—suggesting that his negative intentions were implanted (e.g., the transgressor was abused by his father). We found support for existing blame models by showing that participants blamed more in the purely intentional condition than in the impulsive condition. Crucially, we also found support for our primary hypothesis: Blame responses were significantly attenuated in the implantation condition, compared to the purely intentional condition, even though both conditions presented identical information regarding the actor's intentionality (and participants rated the actor's bullying as equally intentional). Finally, a test of mediation suggested that blame attenuation via implantation information was fully mediated by the perception that the bully lacked free will over becoming a bully. Interestingly, observers still believed that the bully was free to choose to stop bullying others in subsequent interactions. Thus, a specific free will perception—*control of self-formation* (as opposed to *freedom of action*)—seems relevant here. (A further delineation of these perceptions will be provided below.)

Free Will, Determinism, and Ascriptions of Moral Responsibility

Our claim that deterministic (implantation) information reduces blame by diminishing perceptions of free will might suggest that we are taking a stance in the long-running philosophical debate regarding determinism, free will, and moral responsibility. Specifically, our data suggests that people endorse *incompatibilism*: People are not morally responsible if their acts result from deterministic forces. Incompatibilism is typically contrasted with *compatibilism*: People can be held morally responsible even if

their acts result from deterministic forces. We argue, however, that things are more complex than this: People's intuitions regarding free will and moral responsibility are not dichotomous (incompatibilist vs. compatibilist) but reside along a continuum. Prior work however—specifically in experimental philosophy—maintains this distinction, with some studies suggesting that people are incompatibilists, whereas other work suggests that they are compatibilists. We will review that work now.

Nahmias, Morris, Nadelhoffer, & Turner (2005) were pioneers in the empirical exploration of folk intuitions about free will and moral responsibility. They note that philosophers on both sides of the debate claim support for their positions by appealing to putatively general human intuitions about determinism, free will, and blameworthiness. Of course, neither side actually measures these supposed general intuitions, so Nahmias et al. set out to do just that. In two studies, they manipulated a deterministic scenario in two different ways. In one experiment, participants were asked to suppose that a supercomputer can "deduce from the laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time" (Nahmias et al., 2005; p. 566). This includes the behaviors of Jeremy, who, across three different conditions either robs a bank, saves a child, or goes jogging. In spite of being instructed to suppose it was a deterministic universe, in which everything was predictable, the majority of participants (>68%) indicated that Jeremy acts of his own free will across the negative, positive, and neutral scenarios. Also, the majority of participants indicated Jeremy was morally blameworthy (83%) for the robbery and morally praiseworthy (88%) for saving the child. In a second experiment, participants read about a different deterministic scenario, which stated: "The beliefs and values of every person

are caused completely by the combination of one's genes and one's environment" (Nahmias et al., 2005; p. 570). Participants then read about identical twins, separated at birth. One of the twins had been raised by a kindly family and decided to return a lost wallet; the other was raised by a 'jerk' family and decided to keep a lost wallet. Again, despite the information suggesting a deterministic universe, and consistent with the first study, participants indicated that both actors return or keep the wallet of their own free will (76%). The majority also indicated the actors are morally blameworthy or praiseworthy for their actions (>60%). Given that the majority of participants indicate that the actor is morally responsible—despite being in a deterministic universe—these results suggest that folk intuitions are compatibilist.

Nichols & Knobe (2007) built on the work of Nahmias et al. (2005) and presented evidence that observers will sometimes provide compatibilist responses and at other times incompatibilist responses. Specifically, they argued that people are incompatibilists "in the abstract" but become compatibilists when in the throes of a strong negative affective response to a moral transgression. In two experiments, participants read about a deterministic universe, where "everything that happens is completely caused by whatever happened before it," including human decision-making (Nichols & Knobe, 2007; p. 669). Then, in the abstract condition, participants were asked simply whether it is possible for a person to be morally responsible in that universe or, in the concrete condition, whether a man who murders his wife and children is morally responsible for his actions in that universe. In the abstract condition, the majority of participants indicated that a person cannot be morally responsible in a deterministic universe (an incompatibilist response), whereas in the concrete condition, the majority of participants stated that the actor was

morally responsible for killing his family in a deterministic universe (a compatibilist response; Nichols & Knobe, 2007). The authors argue that this divergence is a result of an "affective performance error:" People's true intuitions are incompatibilist, but their affective response to the scenario interferes with the proper application of their beliefs.

Thus, Nichols & Knobe (2007) found that observers would sometimes ignore deterministic information in favor of ascribing moral responsibility. They argued that participants ignored the deterministic information because the moral transgression was affectively arousing. However, in the Gill and Cerce (2013) study (see above), the actor's behavior was emotionally arousing (the actor was a particularly rude and obnoxious jerk), and yet participants *did not* ignore determinism information, and in fact blamed the actor less based on that information. Also, Nahmias et al. (2005) found that participants gave compatibilist responses for positive behaviors and neutral behaviors to the same extent as for moral transgressions. Thus, the "affective performance error" is not well-supported in the literature. Why did Gill and Cerce find divergent results from Nichols & Knobe (2007)? We think this may be because in our study the determinism information is also vivid and emotionally arousing (physical and emotional abuse)—thus, although participants surely felt angry as a result of the actor's transgression, they also likely felt compassion based on the terrible life he had experienced. In contrast, Nichols & Knobe's determinism information was simply an abstract statement that "determinism is true in the universe inhabited by this actor—everything is predictable." Perhaps, then, people's moral judgments are strongly influenced by whatever information is vivid and affectively arousing, and that information can be either determinism information or blameworthiness information.

Feltz, Cokely, and Nadelhoffer (2009) attempted to expand on the work of Nahmias et al. (2005) and Nichols & Knobe (2007). Specifically, the authors argued that not *all* people have the same intuitions regarding determinism and moral responsibility, but rather that there are distinct groups that either consistently apply compatibilist or incompatibilist philosophies across scenarios. Feltz et al. (2009) presented the same vignettes from the Nichols & Knobe (2007) study (regarding a deterministic universe) and presented both the concrete (high affect) and abstract (low affect) scenarios in a within-subjects design. The majority of participants provided matched incompatibilist responses (indicating no one was morally responsible in the deterministic universe) and a smaller percentage provided matched compatibilist responses. Only a small minority (8%) of participants provided mismatched responses, suggesting that most people use consistently compatibilist or incompatibilist theories. Feltz et al. (2009) replicated these findings (in a second study) when they provided the participants with the same vignettes but then asked whether the actor "acts of his own free will" for both the concrete and abstract scenarios. Feltz et al. suggest that the intuitions of most people are consistently incompatibilist and that this is true regardless of affect. They also note, however, that a substantial minority of people have consistent compatibilist intuitions.

In sum, Gill and Cerce's (2013) work on implantation suggests that people have incompatibilist intuitions. The work of Feltz et al. (2009) is largely congruent with this suggestion. On the other hand, Nahmias et al. (2005) find that people are generally compatibilists, holding on to a perception of free will and moral responsibility even when an actor's behavior is fully determined. Nichols and Knobe (2007) corroborate this conclusion, at least when people are responding to concrete, affectively charged

scenarios. The literature is, to put it mildly, maddeningly inconsistent. Yet, we suggest that there is important meaning behind this seeming inconsistency: People do not have either one intuition or the other; they have both.

Compatibilist and Incompatibilist Intuitions Live Side-by-Side Within the Same Mind

We argue that people conceive of free will capacities as residing along a continuum—people view targets in terms of gradations in free will, rather than in terms of free will being "fully present" or "fully absent." As such, they are able to provide what appear to be inconsistent incompatibilist and compatibilist intuitions, when in fact this apparent inconsistency represents degrees of free will perceptions. Therefore, in some scenarios, they have an intuition that deterministic forces diminish moral responsibility while in other scenarios they have the intuition that moral responsibility remains intact even in the presence of deterministic forces. In other words, blaming practices in the presence of deterministic information regarding the formation of an actor's moral character are malleable. Sometimes, as in Gill and Cerce (2013), an actor may be construed as having diminished responsibility based on an unfortunate history. As noted above, this construal is activated by decreased perceptions of a certain kind of free will: *Control over self-formation*—or the perception that the actor did not have control over acquiring the set of immoral predispositions he has today.

On the other hand, this "incompatibilist response" is not, we predict, the only response that will be given when people learn about an actor's unfortunate history. Instead, we anticipate that there are a variety of psychological mechanisms that can lead observers to ignore or discount the information about an unfortunate history, blaming the

actor as if the history information is irrelevant or non-existent. In this Master's Thesis I will examine two potential mechanisms that I predict will weaken the tendency for determinism information—in the form of information about an unfortunate history—to diminish ascriptions of moral responsibility:

(1) *Reminding observers of an actor's capacity for freedom of action*: The observer is reminded that an unfortunate personal history does not diminish a transgressor's freedom of action, and the actor is steadfastly refusing to use that freedom to make more moral choices.

(2) *Activating a strong motivation to blame in observers*: The observer is motivated to blame based on a sense that immorality is spreading and growing within the community (Tetlock et al., 2007). Such a mindset compels observers to focus on punishing transgressors in order to restore moral order. Thus, when in this mindset, observers should be more strongly focused on factors supporting the culpability of the transgressor, not factors that will mitigate ascriptions of moral responsibility.

I predicted that these processes would reduce the extent to which determinism information concerning an unfortunate history defuses blame. In contrast, I predicted that these manipulations—especially manipulations that remind observers of the capacity for freedom of action—should not influence blame responses in scenarios in which the transgressor behaves immorally as the result of a biological or neurological defect (i.e., a brain tumor that makes it *impossible* for the actor to change) because the actor does not have freedom of action, nor is he capable of changing his behavior. Furthermore, the manipulations presented here should have little effect on a transgressor that is already perceived to behave highly intentionally, without any suggestion that he maintains

diminished free will capacities. In this type of scenario, blame should already be high ("at ceiling"), so any additional blame-amplifying manipulations would be futile. Thus, my expectation was that blaming in the unfortunate history/implantation condition would be especially malleable.

What psychological processes underlie this malleability of moral judgments in the presence of information about an unfortunate history? We remind the reader of Gill and Cerce's (2013) finding that a harm-doer with an unfortunate history is seen to have lacked control of self-formation but as having retained the capacity for freedom of action. Thus, when learning about such a harm-doer and his unfortunate history, the observer, perhaps implicitly, might vacillate between two thought processes: "Yes, his horrible life has 'pushed' him toward having deficient moral character…*but*…he is still a human being and human beings are capable of making new choices and changing their ways." To the extent that the second perception is more strongly activated—especially the notion that the actor has this capacity for freedom of action but steadfastly refuses to employ it—information about an unfortunate personal history will have a relatively weak effect in terms of diminishing moral responsibility. In this way, participants' perceptions of free will increase along the continuum—from diminished perceptions to increased perceptions—resulting in increased blame responses.

Additionally, if the observer is motivated to blame transgressors, as a result of being in the prosecutorial mindset, this could result in the observer ignoring the deterministic implications of implantation information. When in the prosecutorial mindset, the goal of the observer is to identify and punish transgressors to restore order in society. Any mitigating factors, like determinism information, will likely be discounted in

order to achieve this goal. Thus, the activation of either of these construals can disrupt the mitigating effect determinism information regarding an unfortunate history has on blame.

The Present Studies

Gill and Cerce (2013) showed that information about a harm-doer's unfortunate personal history can nullify the impact of perceived intentionality on blame. This happens because such information implies a deterministic scenario in which the harm-doer's unfortunate history has deprived him of control over self-formation. As delineated above, however, we expect that people have inconsistent views regarding whether the deterministic influence of an unfortunate personal history truly does diminish moral responsibility. Indeed, we suggested two mechanisms that we predict will weaken the tendency for historical information to lower blame: (a) Reminders of the capacity for freedom of action, and (b) Heightened blame motivation.

Below we present three experiments to test these ideas. All studies are based around the research design used by Gill and Cerce (2013) and cross the manipulation used in that study with another factor which is predicted to selectively moderate moral responsibility judgments (e.g., blame, free will, punishment) within the implantation/unfortunate history condition. Specifically, our research design is always a 3(*transgression narrative*: caused by biological defect, intentional, intentional with implanted character) X 2(*moderator of implantation effect*: enable implantation effect, disable implantation effect). We include the "biological defect" condition to show that our moderators do not influence moral responsibility judgments in cases of "extreme determinism" in which an actor has neither control of self-formation nor freedom of action and, in fact, is scarcely an intentional agent at all (e.g., Tetlock et al., 2007;

Nahmias, Coates, & Kvaran, 2007). Figure 1 shows idealized predictions that are applicable to all our studies.

Studies 1 and 2 focused on manipulating the salience of freedom of action capacities. In Study 1, subjects were presented with information about how, despite many and varied requests to change, the harm-doer persists in his inappropriate behaviors. Thus, he is being asked both gently and not-so-gently to use his capacity for freedom of action, but he simply refuses to do so. We predicted that, as a result of this additional information, observers would give less weight to implantation information and thus show less blame attenuation than in a "pure" (no reminder) implantation condition. Perceptions of freedom of action are the crucial mediator in this model: The actor's refusal to change emphasizes his freedom of action and his unwillingness to utilize it, resulting in increased blaming. This manipulation should have no effect within the biological defect condition because those with biological defects are seen as lacking freedom of action, and thus a refusal to change cannot be construed as a "choice."

In Study 2 participants completed a manipulation that used their own personal experiences to highlight how an individual maintains freedom of action despite prior implantation of bad habits in one's character. Specifically, participants were instructed to write a short narrative about a time they exerted their willpower to overcome a powerful bad habit or predisposition that they had developed as a result of their personal experiences. Based on this exercise, participants should recognize that all people's personalities are at least partly implanted by personal experiences and yet people are still capable of changing themselves for the better going forward—highlighting the capacity

for freedom of action. With these increased perceptions of freedom of action, participants would be less influenced by implantation information and thus blame more.

Study 3 focused on the motivational mechanism for discounting implantation information. Some research suggests that participants are more punitive and driven to lay blame when in the prosecutorial mindset, and that conservatives are, by default, more likely to be in a prosecutorial mindset even in the absence of an experimental manipulation (Tetlock et al., 2007). These findings suggest that the prosecutorial mindset is more accessible to conservatives than liberals and, therefore, conservatives may be more strongly influenced by this type of manipulation (although Tetlock et al. did not find an interaction). In this study we manipulated whether or not participants learned that the company the transgressor works for is "going to hell in a hand basket" as a result of a steep hike in the amount of office bullying and unresolved harassment claims. We predicted that reading about the company's downward spiral would increase punitiveness, resulting in discounting implantation information and thus greater ascriptions of moral responsibility than in the "pure" implantation condition. Furthermore, we predicted that the effectiveness of this manipulation might vary as a function of political ideology. Finally, we predicted that increased punitiveness in the implantation condition could likely occur via two processes: The motivation to blame could lead to an adjustment of free will perceptions to justify the increased blame response or that there would be a "direct effect" of motivation (that did not result in differences in free will perceptions). Whether motivation increases blame via one process or another is not a crucial prediction for us. Instead, we are more concerned with moral

responsibility ascriptions and especially the possibility that they are particularly malleable in the implantation condition when blame motivation is activated.

Study 1: Refusal to Change

Method

Participants. Participants were recruited from Amazon's Mechanical Turk and paid $0.50 for their participation. Data were collected on 206 participants (67 male) and 12 participants were excluded because they did not complete the study or failed manipulation checks. The average age of participants was 39.82 (SD = 15.02).

Procedure. The experiment was a 3 x 2 between-subjects design. As in Gill and Cerce (2013) participants were randomly assigned to one of three *transgression narrative* conditions that offered different narratives concerning the background of a moral transgression: *biological defect*, *intentional*, or *intentional with implantation* (hereafter called *implantation*). All participants read about James, "who is the Vice President of a moderately sized company who will often humiliate and harass members of his staff." In the biological defect condition, participants learned that James' bullying is a result of a brain tumor in his frontal lobe that increases his aggressiveness and eliminates his ability to inhibit his behavior. In the intentional condition, participants learned that James' bullying is highly intentional (e.g. James has a desire to bully others; even after deliberation he decides bullying is an acceptable behavior). In the implantation condition, participants learned that James' bullying is highly intentional but also that his inclination toward bullying was implanted as a result of an abusive relationship with his father. See Appendix A for all vignettes.

Embedded within roughly half of the James vignettes in each condition was information about various ways in which James' friends and colleagues have attempted to convince James to change his behavior (i.e., "*James, life is difficult enough without us being unkind to each other. Would you please think about what you are doing, think about other people's feelings, and do your best to mellow out?"*). Crucially, despite these pleas, James refuses to change. In the *no refusal to change* condition, participants received no information beyond that in the basic transgression narrative (described above). This *refusal to change* manipulation is meant to remind participants that James can make free choices to change his ways—he simply refuses to do so. See Appendix B for manipulation.

Measurement of Blame, Free Will, and Other Relevant Constructs. All dependent variable items were measured on a 1 (strongly disagree) to 5 (strongly agree) scale (See Appendix C). After reading the vignette and the refusal to change manipulation, participants completed items tapping our two primary dependent variables, blame (e.g., *I blame James for his bullying*) and free will. The four blame items were aggregated to create an index of blame ($M = 3.73$, $SD = 1.06$, $\alpha = 0.921$). Seven items tapped perceived freedom of action (e.g., *Using his capacity for free will, James could choose to STOP being a bully*), which were also aggregated ($M = 3.65$, $SD = 1.21$, $\alpha = 0.900$). Five items tapped perceived control of self-formation (e.g., *James had free will in terms of BECOMING a bully*), which were aggregated ($M = 2.76$, $SD = 0.98$, $\alpha = 0.980$). A factor analysis confirmed that freedom of action and control of self-formation are distinct free will perceptions: Freedom of action items loaded onto one factor at 0.87 or higher (eigenvalue = 7.63; 63.58% of variance explained) and control of self-formation

items loaded onto a second factor at 0.69 or higher (eigenvalue = 2.29; 19.09% of variance explained).

Then participants completed four manipulation check items to assess the success of our *refusal to change* manipulation (e.g., *James has free will—he just needs to USE it*; $M = 3.45$, $SD = 1.23$, $\alpha = 0.951$). The manipulation check items were meant to be distinct from the freedom of action items in one critical way: While both sets of items tap perceptions that James has freedom of action, the manipulation check items are intended to highlight that he is *not using that capacity*. Therefore, participants would have to recognize that he both maintains freedom of action and that he is not applying that capacity—which should be particularly apparent after the refusal to change manipulation.

Finally, participants responded to two different types of punishment items. Four items tapped participants' preference for punishment that would attempt to encourage James to change via the threat of various sanctions (e.g., *James should receive a clear warning from a superior: You will not be eligible for a raise until you start treating others with respect*). These punishment items were aggregated to create an "encourage him to change" punishment index ($M = 3.99$, $SD = 0.91$, $\alpha = 0.883$). A separate set of five punishment items measured participants' preference for types of punishment that would "mess with" James and make his life difficult (e.g., *It would be great to learn that one of James' coworkers "accidently" spilled water on his laptop, destroying its contents*), which were aggregated to create a "mess with him" punishment index ($M = 2.35$, $SD = 0.99$, $\alpha = 0.919$). A factor analysis confirms that these punishment responses are distinct constructs, with the "encourage him to change" punishment items loading onto one factor at 0.65 or higher (eigenvalue = 4.50; 50.04% of variance explained) and "mess with him"

punishment responses loading onto a second factor at 0.71 or higher (eigenvalue = 2.48; 27.60% of variance explained).

Based on previous studies (Gill & Cerce, 2013), the intentional and implantation vignettes had already been equated on perceptions of intentionality and negative behavior expectations, so those assessments were not completed here[2]. Furthermore, we were not concerned with perceptions of suffering, as our findings in Gill & Cerce (2013) provide evidence that this perception does not mediate the effect the implantation manipulation has on blame judgments: Even though perceptions of suffering were significantly higher in the implantation condition, it was diminished control of self-formation—not suffering perceptions—that resulted in mitigated blame responses[3].

---

2 In a previous study, using the same vignettes presented here (regarding James' bullying), participants completed items tapping perceptions of intentionality (e.g., *James looks for opportunities to embarrass or harass his staff members*) and negative behavior expectations (e.g., *James' staff has good reason to avoid him during the work day*). Participant responses were both high and nearly equal in the intentional ($M = 4.57$) and implantation ($M = 4.47$) condition ($t < 1$) and both were significantly greater than in the biological defect condition [$M = 2.90$; $t\ (120) = 10.82$, $p < 0.001$; $t\ (120) = 10.08$, $p < 0.001$]. Negative behavior expectations were not significantly different across the conditions (ts < 1.5). Therefore, results suggest that perceptions of intentionality and negative behavior expectations are equated across the intentional and implantation conditions.

3 This finding may be somewhat surprising given the work of Gill, Andreychik, and Getty (2013), which suggests that external explanations reduce compassion via perceptions of suffering. However, the results presented in their Study 3 suggest that this mediation effect is moderated by identification with the target of suffering: The more strongly the observer identifies with the target group, the stronger the relationship between perceived suffering and compassion. In contrast, in the manipulation presented here, James is an office executive who harasses his employees—thus it is unlikely that observers strongly identify with him. We think it is this lack of identification that results in perceived suffering not strongly predicting blame responses.

Results

To test our predictions, we conducted a two-way independent ANOVA with *transgression narrative* (biological defect, intentional act, intentional act with implantation) and *refusal to change* (no, yes) predicting our dependent variables (all analyses are based on the same ANOVA model unless otherwise noted).

First, we tested whether our refusal to change manipulation was effective and found a marginal main effect of refusal to change manipulation on perceptions of James' willingness to try to change his behavior, $F(1, 188) = 3.66$, $p = 0.057$. When participants were provided information suggesting that James has had opportunities to change, but refused to do so, perceptions that James was not trying were significantly higher ($M = 3.57$) than when participants were not provided this additional information ($M = 3.36$). Upon further examination of the means, a focused condition-by-condition analysis showed that the refusal to change manipulation only affected the manipulation check items in the intentional condition: Perceptions of James' refusal to change were significantly higher when the refusal to change manipulation was present ($M = 4.52$) versus absent ($M = 4.14$), $t(188) = 2.00$, $p = 0.047$, $d = 0.29$. There were no significant differences in the biological defect or implantation conditions (ts < 1). These findings suggest that people are construing the refusal to change manipulation very differently across our conditions. Specifically, the manipulation appears to amplify existing beliefs that James is not trying in the intentional condition, but does not affect this perception in the implantation or biological defect condition. We had anticipated that the refusal to change manipulation would be construed similarly across the intentional and implantation conditions (conditions in which—based on the results of Gill & Cerce—freedom of

action is thought to be unimpaired), but differently in the biological defect condition (where freedom of action is not possible). This assumption is not supported by our data.

We proceeded to test our primary predictions because, of course, our manipulation check items could be flawed and painting a false image of the inadequacy of our refusal to change manipulation. Our primary predictions were about blame responses. We found a significant main effect of transgression narrative, $F(2, 188) = 119.07$, $p < 0.001$. In this and all subsequent studies, we replicated existing blame literature. Thus, we found that blaming was higher in the intentional condition ($M = 4.52$) than in the biological defect condition ($M = 2.66$), $t(188) = 14.72$, $p < 0.001$, $d = 2.14$. Also, we replicated Gill and Cerce (2013): Blame was significantly lower in the implantation condition ($M = 4.04$) than in the intentional condition, $t(188) = 3.76$, $p < 0.001$, $d = 0.54$. However we did not find a main effect of the refusal to change manipulation or the predicted interaction (Fs < 1.3, ps > 0.252). Therefore, we did not confirm our primary predictions regarding blame responses: that blame in the implantation condition would be significantly higher when the refusal to change manipulation is present rather than absent (See Figure 2 for pattern of responses). Beyond this primary analysis, we examined whether the effects of our manipulations were moderated by Social Explanatory Style, Need for Cognitive Closure, or General Beliefs about Free Will. We found no evidence of moderation.

Next, we turned to analysis of free will perceptions. We found a significant main effect of transgression narrative, $F(2, 188) = 91.36$, $p < 0.001$. Pairwise comparisons revealed that freedom of action perceptions were significantly higher in the intentional condition ($M = 4.48$) compared to the implantation condition ($M = 4.22$), $t(188) = 1.93$, p

= 0.054, d = 0.28, and the biological defect condition (M = 2.30), $t(188) = 16.71$, $p <$ 0.001, d = 2.43. Furthermore, perceptions were significantly higher in the implantation condition versus the biological defect condition, $t(188) = 15.11$, $p < 0.001$, d = 2.20. Thus, we did not replicate findings from Gill & Cerce (2013) that perceptions of freedom of action were not significantly different in the intentional and implantation conditions. We did, however, replicate our existing finding that freedom of action perceptions were significantly lower in the biological defect condition. The main effect of the refusal to change manipulation and the interaction were not significant (Fs < 1; ps > 0.399).

Despite the ineffectiveness of the refusal to change manipulation on blame responses and freedom of action perceptions, we did find significant main effect of both the transgression narrative, $F(2, 188) = 31.36$, $p < 0.001$, and the refusal to change manipulation, $F(1, 188) = 4.27$, $p = 0.040$, on perceived control of self-formation. Replicating Gill and Cerce (2013), pairwise comparisons revealed that participants made significantly decreased ascriptions of freedom of self-formation in the implantation condition (M = 2.64) as compared to the intentional condition (M = 3.44), $t(188) = -5.30$, $p < 0.001$, d = 0.77. Additionally, perceptions of control of self-formation were significantly lower in the biological defect condition (M = 2.26) compared to both the implantation condition, $t(188) = -2.50$, $p = 0.013$, d = 0.36, and intentional condition, $t(188) = -7.76$, $p < 0.001$, d = 1.13. Results involving the refusal to change manipulation suggested that control of self-formation perceptions were significantly higher in the refusal to change condition (M = 2.91) as compared to the no refusal to change condition (M = 2.65), $t(188) = 2.04$, $p = 0.040$, d = 0.29. Also, interestingly, although the interaction was not significant, inspection of means suggested that the effect of the

refusal to change manipulation on perceived control of self-formation was strongest in the implantation condition (See Figure 3 for pattern of means). Indeed, the effect of the refusal to change manipulation was not significant in the biological defect condition and the intentional condition (ts < 1; ps > 0.322). In the implantation condition, in contrast, control of self-formation ascriptions significantly increased when participants were reminded of James' opportunities to change (M = 2.84) versus when they were not provided this additional information (M = 2.43), $t(188) = 1.95$, $p = 0.052$, $d = 0.28$. Thus, the deterministic (implantation) information regarding James' history appears to be a less compelling explanation for James' bad behavior when participants are made aware that many people have put pressure on James to change, but he has refused to do so.

As for punishment preferences, there was a main effect of transgression narrative on "encourage him to change" punishment responses, $F(2, 188) = 52.62$, $p < 0.001$. Pairwise comparisons revealed that this type of punishment was rated as equally desirable in both the intentional (M = 4.38) and implantation conditions (M = 4.3; $t < 1$). However, "encourage him to change" punishment was significantly less preferred in the biological defect condition (M = 3.22) compared to both the intentional, $t(188) = -8.80$, $p < 0.001$, $d = 1.28$, and implantation conditions, $t(188) = -8.90$, $p < 0.001$, $d = 1.29$. The main effect of refusal to change manipulation was not significant ($F < 1$; $p > 0.747$). However, there was a marginal interaction of transgression narrative and the refusal to change manipulation, $F(2, 188) = 2.482$, $p = 0.086$ (See Figure 4). Pairwise comparisons did not reveal any significant or marginal effects of refusal to change manipulation within each condition. There was, however, a weak trend such that refusal to change increases these punishment judgments in the intentional condition (M = 4.52) versus when the refusal to

change manipulation is absent ($M = 4.25$), $t(188) = 1.43$, $p = 0.166$, $d = 0.20$. In the implantation condition there was a weak trend toward reversal of this pattern: Attempts to change James through punishment are *less* preferred when the refusal to change manipulation is present ($M = 4.22$) than absent ($M = 4.51$), $t(188) = -1.60$, $p = 0.110$, $d = 0.23$. There is no effect of the manipulation ($t < 1$; $p > 0.465$) in the biological defect condition, which supports our prediction that free will relevant manipulations should not influence responses there. In other words, participants are responding to the refusal to change manipulation differently depending on which transgression narrative condition they are in. In the intentional condition, the refusal to change manipulation seems to amplify existing punishment responses. In contrast, in the implantation condition it appears that participants recognize that although James maintains some freedom of action, he does not take advantage of opportunities to change, therefore punishment responses aimed at changing his behavior will be ineffective.

As for "mess with him" punishment items, there was a main effect of condition, $F(2, 188) = 17.186$, $p < 0.001$. The main effect of refusal to change manipulation and the interaction were not significant (Fs < 1; ps > 0.342). Pairwise comparisons revealed that participants preferred "mess with him" punishment items more in the intentional condition ($M = 2.82$) compared to the implantation condition ($M = 2.40$), $t(188) = 2.50$, $p = 0.013$, $d = 0.36$, and the biological defect condition ($M = 1.86$), $t(188) = 5.81$, $p < 0.001$, $d = 0.84$. Furthermore, participants preferred this form of punishment significantly more in the implantation condition, as compared to the biological defect condition, $t(188) = 3.37$, $p = 0.001$, $d = 0.49$. Unsurprisingly, the more intentional James' behavior is, the more willing participants are to condone punishment responses that "mess with him."

Discussion

In Experiment 1, transgression narratives were crossed with a manipulation in which some participants learned that James' co-workers have done their best to persuade him to change but he refuses, whereas other participants heard of no such persuasion efforts. Our prediction was that this "refusal to change" manipulation would selectively increase blame and other indicators of moral responsibility in the implantation condition. We did not find our predicted results: Blame responses in the implantation condition (and the other transgression narrative conditions) were completely unaffected by the refusal to change manipulation.

We also examined whether free will perceptions and punishment responses were influenced by the refusal to change manipulation. As a result of the refusal to change manipulation, participants indicated increased perceptions of James' control over his self-formation. This difference appears to manifest most in the implantation condition, where information suggesting that James has not taken advantage of opportunities to change indicate to participants that he had greater control over his self-formation than the deterministic information regarding his history would otherwise suggest. Interestingly, these results suggest that the refusal to change manipulation is 'undoing' the mechanism (diminished control of self-formation) that drives the implantation effect found in Gill & Cerce (2013). As for perceived freedom of action, the refusal to change manipulation did not influence these perceptions either within or across transgression narrative conditions. Our goal when creating the refusal to change manipulation was to selectively activate freedom of action perceptions. The results of the free will perceptions suggest that we did something other than what we wanted: Namely, we did not manipulate freedom of action

perceptions but rather control of self-formation perceptions. Instead of activating the perception that James had the capacity to overcome his predispositions, our refusal to change manipulation instead increased perceptions that he had greater control over his personality development.

As for punishment responses, the evidence weakly suggested that the refusal to change manipulation influenced "encourage James to change" punishment responses differently in the intentional and implantation conditions. The refusal to change manipulation results in increased punishment judgments in the intentional condition and decreased punishment in the implantation condition. While this reversal is somewhat confusing (and weak enough as to not warrant extended interpretation), it seems plausible that attempts to get James to change via punishment are less preferred in the implantation condition because he has not taken advantage of these opportunities in the past, so future attempts will be unsuccessful.

The question remains as to why we did not find stronger effects on blame and freedom of action perceptions. Our manipulation checks revealed that participants do recognize that James has opportunities to change and he just chooses not to *use* them—but only in the intentional condition. Therefore, it appears that we failed to manipulate what we wanted to manipulate: Increasing freedom of action perceptions, specifically in the implantation condition. We think this might be the case because, although the refusal to change manipulation may suggest that James chooses not to use his opportunities to change, it makes no claims as to *why* he does not use these opportunities. Indeed, it may be the case that the manipulation may reinforce that James cannot change—or else clearly he would have done so given so many opportunities. In other words, as a result of

the refusal to change manipulation, James is seen as having greater control of his self-formation, but not greater freedom of action. This is an interesting composition of free will perceptions that needs to be examined further. Specifically, at first blush, it would appear that a manipulation that increases perceptions of control of self-formation should necessarily increase perceptions of freedom of action—because it is through exerting agentic capacities that individuals are able to control the formation of their character. However, this is not necessarily the case (as the results of this study suggest). Instead, while the refusal to change manipulation suggests that James maintains some control of his self-formation (by refusing opportunities to change), he may not necessarily 'freely' refuse these opportunities. Although James has contributed to his own self-formation (by refusing opportunities to change), he still did not make a "free choice" to develop these negative intentions. This may explain why our blame responses are not as robust as predicted in the implantation condition.

As a result of this experiment, future directions should be aimed at examining to what extent manipulations that heighten perceptions of freedom of action versus control of self-formation differentially influence blame ascriptions. The results of this study suggest that manipulations that selectively heighten control of self-formation perceptions may not amplify blame responses very robustly.

Study 2: Highlighting the Ability of the Self to Overcome Predispositions

Method

Participants. Participants were 148 Lehigh University undergraduates (68 male) who participated for course credit. The average age was 19.35 (*SD* = 1.35).

Procedure. The experiment was a 3 (transgression narrative: *biological defect, intentional*, or *implantation*) x 2 (overcoming predispositions task vs. control task) between-subjects design. Participants read the same vignettes about James presented in Study 1 (Appendix A).

Before reading about James' bullying, participants either were randomly assigned to the overcoming predispositions manipulation or a control task. The cover story indicated that this task was a separate study from what was to follow. Participants in the overcoming predispositions condition were instructed to "think of a time when you overcame—at least for a while—a powerful bad habit in your life" (See Appendix D for complete instructions). The manipulation is meant to highlight how all people are partly shaped by previous life experiences, yet are still able to change going forward and exert self-regulatory control over the actions. Participants were asked about a time when they *overcame* a challenging predisposition, which should highlight freedom of action—and suggest that if a person puts in enough effort they can choose the right course of action and exercise self-regulatory efforts (including James). In contrast, participants in the control task were asked to "think of a time when you had free time and had to make a decision about how to fill it." This task was equated in terms of completing a personal narrative task without prompting participants to think of concepts relevant to freedom of action perceptions.

Measurement of Blame, Free Will, and Other Relevant Constructs. After reading the transgression narrative vignette and completing the overcoming predispositions manipulation, participants completed all of the dependent variable items (same items as in Study 1, provided in Appendix C).

Results

Our primary predictions were about blame. We found a significant main effect of transgression narrative, $F(2, 142) = 65.66$, $p < 0.001$. Blame was significantly higher in the intentional condition ($M = 4.03$) compared to the implantation ($M = 3.26$), $t(142) = 5.64$, $p < 0.001$, $d = 0.94$, and biological defect conditions ($M = 2.46$), $t(142) = 11.48$, $p < 0.001$, $d = 1.92$. Blame was also significantly higher in the implantation condition versus the biological defect condition, $t(142) = 5.83$, $p < 0.001$, $d = 0.92$. The main effect of the overcoming predispositions manipulation was weak, $F(2, 142) = 2.09$, $p = 0.150$. The pattern of blame responses was opposite of what might be expected: Participants blamed more in the control task ($M = 3.37$) compared to the overcoming predispositions task ($M = 3.31$). See Figure 5 for pattern. The interaction of transgression narrative and overcoming predispositions manipulation was not significant ($F < 1$; $p > .402$). In order to test our predicted effect further, we conducted within condition contrasts. There were no significant effects of the overcoming predispositions task within any of the transgression narrative conditions (Fs < 1.75; ps > .19).

As for freedom of action responses: There was a significant main effect of transgression narrative condition, $F(2, 142) = 73.18$, $p < 0.001$. Pairwise comparisons revealed that perceptions were significantly higher in the intentional condition ($M = 4.22$) and the implantation condition ($M = 4.13$) compared to the biological defect condition ($M = 2.79$), $t(142) = 10.75$, $p < 0.001$, $d = 1.80$; $t(142) = 10.19$, $p < 0.001$, $d = 1.71$. Replicating Gill & Cerce (2013), freedom of action perceptions were not significantly different in intentional and implanted conditions ($t < 1$; $p > 0.474$). The main effect of overcoming predispositions manipulation was not significant, ($F < 1.8$; $p > 1.74$). There

was a significant interaction of transgression narrative and overcoming predispositions manipulation predicting freedom of action, $F(2, 142) = 3.24$, $p = 0.042$. See Figure 6 for the pattern of means. Pairwise comparisons revealed that, in the biological defect condition, freedom of action perceptions were significantly lower when participants completed the overcoming predispositions task ($M = 2.53$) versus the control task ($M = 3.06$), $t(142) = -2.85$, $p = 0.005$, $d = 0.47$. Perceptions of freedom of action within the intentional and implantation conditions were not significantly different when completing the overcoming predispositions versus control task (ts < 1; ps > 0.689).

Next we analyzed control of self-formation. There was a significant main effect of transgression narrative, $F(2, 141) = 13.61$, $p < 0.001$. Perceptions were significantly greater in the intentional condition ($M = 2.94$) than the implantation ($M = 2.29$), $t(141) = 4.57$, $p < 0.001$, $d = 0.76$, and biological defect conditions ($M = 2.30$), $t(141) = 4.49$, $p < 0.001$, $d = 0.75$. Perceptions were not significantly different between the implantation and biological defect conditions ($t < 1$; $p > 0.963$). There was also a main effect of overcoming predispositions manipulation, $F(1, 141) = 5.11$, $p = 0.025$. Participants who completed the overcoming predispositions task indicated decreased perceptions of control of self-formation ($M = 2.38$) than participants who completed the control task ($M = 2.64$). Given that the overcoming predispositions task reminded participants that they have acquired personality traits and predispositions from their personal history, regardless of any intention on their part to do so, it is not surprising that this task resulted in decreased perceptions of self-formation. Importantly, however, it was not our intention to alter perceptions of self-formation via this manipulation. It seems likely that this unintended effect is explaining the weak reversal in blame judgments we found—that participants

blame more in the control task versus the overcoming predispositions task. Although the interaction of transgression narrative and overcoming predispositions manipulation was not significant ($F < 1$; $p > 0.789$), an examination of the means suggests that the main effect of overcoming predispositions manipulation may be driven by differences in perceptions in the intentional condition: In that condition, control of self-formation perceptions were reduced in the overcoming predispositions condition ($M = 2.75$) as compared to the control condition ($M = 3.13$), $t(141) = 1.84$, $p = 0.067$, $d = 0.30$. Perceptions were not significantly different in the implantation or biological defect conditions (ts < 1.1; ps > 0.263). These results suggest that our manipulation must have made clear to participants that it is *hard* to overcome predispositions, resulting in an activation of diminished control of self-formation perceptions rather than increased freedom of action perceptions.

Next, we examined punishment judgments. Looking at "encourage him to change" punishments, there was a significant main effect of transgression narrative, $F(2, 142) = 5.79$, $p = 0.004$. Pairwise comparisons revealed that "encourage him to change" punishment responses were significantly higher in the intentional ($M = 4.07$) and implantation conditions ($M = 4.03$) compared to the biological defect condition ($M = 3.60$), $t(142) = 3.09$, $p = 0.002$, $d = 0.51$; $t(142) = 2.79$, $p = 0.006$, $d = 0.46$. Punishment responses were not significantly different in the intentional and implantation conditions ($t < 1$; $p > 0.760$). The main effect of overcoming predispositions manipulation was marginal, $F(1, 142) = 2.17$, $p = 0.142$. There was a significant interaction of transgression narrative and overcoming predispositions manipulation, $F(2, 142) = 8.54$, $p < 0.001$. See Figure 7 for pattern. Pairwise comparisons revealed that, in the implantation condition,

"encouraging him to change" punishment responses were significantly higher when the overcoming predispositions manipulation was presented ($M = 4.30$) versus the control task ($M = 3.76$), $t(142) = 2.51$, $p = 0.013$, $d = 0.46$. This pattern of findings is in line with our thinking: Given the reminder that freedom of action remains intact despite a lack of control over self-formation (e.g., "*I* have managed to transcend a bad habit that was implanted in *me*."), deterministic information regarding James' history was less compelling, resulting in increased punishment responses. Interestingly, the pattern of responses was the opposite for the biological defect condition: "encourage him to change" punishment responses were significantly lower when participants completed the overcoming predispositions task ($M = 3.30$) versus the control task ($M = 3.90$), $t(142) = -2.79$, $p = 0.006$, $d = 0.46$. Given that this manipulation was meant to highlight a person's ability to overcome predispositions and change for the better, it is not surprising that being reminded of this capacity reduces punishment in the biological defect condition. The manipulation is a reminder of a capacity that James cannot possibly have because of his brain tumor. These results are consistent with freedom of action perceptions in the biological defect condition: being reminded of one's own capacity for freedom of action facilitated participants' recognition that James would never have this capacity (due to his brain tumor), therefore he has diminished freedom of action and should be punished less. However, punishment responses showed a similar pattern in the intentional condition: Punishment was higher when participants completed the control task ($M = 4.32$) versus the overcoming predispositions task ($M = 3.83$), $t(142) = 2.24$, $p = 0.027$, $d = 0.37$. In sum, it is clear that participants are responding to the overcoming predispositions manipulation differently depending on which transgression narrative condition they are

in: The manipulation amplified punishment responses in some conditions (implantation) but not others (biological defect and intentional).

Finally, we analyzed the "mess with him" punishment responses. There was a significant main effect of transgression narrative, $F(1, 142) = 3.92$, $p = 0.022$. Participants indicated increased preference for this type of punishment in the intentional condition ($M = 2.21$) versus the biological defect condition ($M = 1.79$), $t(142) = 2.80$, $p = 0.006$, $d = 0.46$. "Mess with him" punishment responses were not significantly different in the implantation condition ($M = 2.01$) compared to either the intentional or biological defect conditions (ts < 1.44; ps > 0.152). Again, not surprisingly, participants are most willing to condone "mess with him" forms of punishment the more it is clear that James' behavior is intentional with no determinism information presented. The main effect of the overcoming predispositions manipulation and the interaction were not significant (Fs < 1; ps > 0.415).

Discussion

In Experiment 2, we crossed our transgression narrative manipulation with a manipulation of whether or not participants were reminded of their own past success at using willpower to change an implanted bad habit. Prior to reading about James, participants completed a task under the guise of a separate study. During this task, participants were either told to think of a time when they were able to overcome a strong predisposition they formed as a result of their personal experiences or a matched control task. We predicted that the overcoming predisposition task would increase perceptions of freedom of action and increase blame responses, specifically in the implantation condition.

Results suggest that participants were significantly more likely to espouse "encourage him to change" punishment for James in the implantation condition, after completing the overcoming predispositions task. Given the overcoming predispositions task, the deterministic personal history information was seen as a less compelling reason to mitigate punishment responses. In contrast, in the biological defect and intentional conditions, punishment responses were significantly reduced by the overcoming predispositions task as compared to the control task—suggesting that this manipulation is not influencing punishment responses similarly across transgression narratives.

Why would the overcoming predispositions task *decrease* punishment in the biological defect and intentional conditions? In the biological defect condition, perceptions of freedom of action were significantly reduced in the overcoming predispositions condition as compared to the control condition. It seems likely that, in the biological defect condition these perceptions are driving punishment responses. In the biological defect condition, James is incapable of changing his behavior—as such, highlighting one's own capacity to change may result in placing James capabilities in stark contrast to this recognition. Therefore, perceptions of James' freedom of action decrease, as do punishment responses. As for the intentional condition, decreased punishment responses after the overcoming t task may be a result of the decreased perceptions of control of self-formation—as results suggest that the overcoming predispositions manipulation increased control of self-formation perceptions most in this condition. Therefore, in the biological defects and intentional conditions, the overcoming predispositions manipulation results in decreased free will perceptions, which are reflected in decreased punishment responses. In contrast, in the implantation condition,

the overcoming predispositions manipulation had no effect on either freedom of action or control of self-formation perceptions—yet, it increased "encourage him to change" punishment responses.

Future replications of this study should be aimed at exploring why the overcoming predispositions manipulation influenced the transgression narrative conditions differently. It seems to be the case that in the biological defect and intentional conditions, the manipulation results in participants' diminished free will perceptions. In these conditions, the overcoming predispositions manipulation may have prompted participants to consider how they have experienced diminished control over self-formation or freedom of action in their own lives. In implantation condition, the manipulation increases punitiveness, but with no comparable alterations to free will perceptions. Therefore, it may be the case that this manipulation is altering an unmeasured third variable, which is subsequently resulting in increased punishment responses. Some possibilities for this unmeasured variable could be blame-proneness—which is an individual difference measure that determines he ease and readiness by which people are willing to make blame ascriptions. Those higher in blame proneness may be more likely to pick up on and respond to aspects of a judgment context that allow them to blame more (i.e., additional information that allows for one to selectively ignore deterministic information in favor of ascribing moral responsibility). As such, the influence of other potential unmeasured variables, whether moderators (such as blame proneness) or mediators, should be explored.

An additional question of interest is why we did not find our predicted blame responses in this study. We found marginal increases in blame after completing the

control task in the biological defect and intentional conditions. Similar to Experiment 1, the overcoming predispositions manipulation altered control of self-formation perceptions—although in this study, it resulted in decreased perceptions. Additionally, the results suggest that we did not increase perceptions of freedom of action. Therefore, the key issue here is that we need to better develop manipulations that can successfully and robustly remind participants of freedom of action capacities. It may also be the case that reminding participants that they have formed predispositions as a result of a formative personal history could have stuck with them longer than the reminder that they actually *overcame* these predispositions—and it was the former construal that influenced blame responses. As such, responses for the overcoming predispositions task (which were free response) should be coded in terms of whether their focus was on their personal success or failure in overcoming predisposition, to determine whether this mediated results.

Study 3: Prosecutorial Mindset Manipulation

Method

Participants. We recruited 180 participants (86 males) from Amazon's Mechanical Turk and they were paid $0.50 for their participation. The average age of participants was 37.64 (*SD* = 13.52). Thirteen participants were excluded from analyses either because they did not complete the study or failed manipulation checks.

Procedure. The experiment is a 3 (transgression narrative condition: *biological defect*, *intentional*, or *implantation*) x 2 (prosecutorial mindset manipulation: prosecutorial mindset or generic information), between-subjects design. Participants read the same vignettes about James presented in Study 1 (Appendix A).

Prior to reading about James, participants either were randomly assigned to read a passage about James' company that described how office respect has rapidly declined over the past couple of years (prosecutorial mindset condition) or about how manufacturing at the company has changed over the past couple of years (generic information condition). The prosecutorial mindset manipulation was modified based on manipulations created by Tetlock and colleagues (2007) whereby it seems like society (i.e., James' company) is "going to hell in a hand basket" and that transgressors aren't being punished for their violations. This manipulation has been shown to motivate punitiveness in the service of restoring social order. The generic information condition simply described how manufacturing at the company has changed from an assembly line to mainly office work. See Appendix E for manipulation.

Measurement of Blame, Free Will, and Other Relevant Constructs. After reading the prosecutorial mindset manipulation and the transgression narrative vignette, participants completed all of the dependent variable items (same items as in Study 1, provided in Appendix C).

Results

Our primary predictions pertain to blaming. With transgression narrative condition and prosecutorial mindset manipulation predicting blame, there was a significant main effect of transgression narrative which replicated our existing pattern of findings, $F(2, 161) = 60.63$, $p < 0.001$. The main effect of prosecutorial mindset manipulation and the predicted interaction were not significant (Fs < 1). However, as discussed above, Tetlock & colleagues (2007) have found that the effectiveness of the prosecutorial mindset manipulation varies based on political ideology. Therefore, next we

looked at transgression narrative and prosecutorial mindset manipulation predicting blame, moderated by political ideology. Political ideology was measured on a 1 to 7 Likert scale (from very liberal to very conservative).

We found a significant interaction of transgression narrative, prosecutorial mindset manipulation, and ideology, $F(2, 155) = 11.57$, $p < 0.001$. To illuminate this interaction, we computed predicted means separately for liberals (-1 SD on the ideology measure) and conservatives (+1 SD on the ideology measure) For liberals, there was a marginal effect of mindset in the implantation condition: Liberals blamed more when in the prosecutorial mindset ($M = 4.24$) than in the generic condition ($M = 3.75$) $t(155) = 1.78$, $p = 0.076$, $d = 0.28$. See Figure 8. Therefore we found our predicted effect for liberals: The prosecutorial mindset manipulation led them to discount deterministic information regarding James' history and to blame more. Conversely, in the biological defect condition this pattern was reversed: Liberals blamed more in the generic condition ($M = 3.15$) than in the prosecutorial mindset ($M = 2.38$), $t(155) = 2.81$, $p = 0.005$, $d = 0.45$. Although this reversal was not predicted, it is still in line with our thinking about differences in people's responses to biological versus historical determinism: While liberals in a prosecutorial mindset are willing to discount deterministic information concerning a harm-doer's *history*, information about biological determinism remains compelling to them even when they are in a prosecutorial mindset. In the biological defect condition (where the prosecutorial mindset reduced blame for liberals, it seems likely that liberals see the company culture as a further deterministic explanation for why James behaves poorly: A guy with a brain tumor might be more easily "swept away" by the prevailing disturbed cultural norms at the company. For liberals, blame responses

within the intentional condition were unaffected by the prosecutorial mindset manipulation ($t < 1$; $p > 0.427$).

As for conservatives, pairwise comparisons revealed that, in the biological defect condition, conservatives blamed significantly more in the prosecutorial mindset ($M = 3.79$) than in the generic condition ($M = 2.69$), $t(155) = 3.66$, $p < 0.001$, $d = 0.58$. See Figure 9. Although we had expected that the prosecutorial mindset would not matter in the biological defect condition (replicating Tetlock et al., 2007), this result is in line with the general idea that motivation to blame can lead to the discounting of determinism information (although it is inconsistent with our expectation that this will happen more for historical rather than biological determinism). Specifically, the biological defect condition is a kind of deterministic scenario (i.e., James has a brain tumor that he does not choose to have and cannot change). Conservatives appear to discount the blame-reducing implications of that information when we put them in a prosecutorial mindset. For conservatives, blame responses in the intentional and implantation conditions were unaffected by the prosecutorial mindset manipulation (ts < 1; ps > 0.416).

We wondered whether the prosecutorial mindset manipulation resulted in an adjustment of free will perceptions or whether there was a "direct effect" of motivation on blame. For perceptions of freedom of action, there was a significant main effect of transgression narrative, $F(2, 155) = 33.96$, $p < 0.001$. Perceptions were significantly lower in the biological defect condition ($M = 2.75$) versus the implantation ($M = 4.37$), $t(155) = 11.90$, $p < 0.001$, $d = 1.91$, and intentional conditions ($M = 4.55$), $t(155) = 13.28$, $p < 0.001$, $d = 2.13$. Perceptions were not significantly different in the intentional and implantation conditions ($t < 1.3$; $p > 0.193$). There were also significant interactions of

transgression narrative by motivation manipulation, $F(2, 155) = 5.198$, $p = 0.007$, and transgression narrative by ideology, $F(2, 155) = 3.83$, $p = 0.024$. These interactions were qualified by a significant three-way interaction of transgression narrative, prosecutorial mindset manipulation, and ideology, $F(2, 155) = 8.10$, $p < 0.001$ (See Figures 10 and 11 for the pattern of means computed separately for liberals and conservatives). First, we will consider liberals. Pairwise comparisons revealed that in the biological defect condition, liberals indicated lowered perceptions of freedom of action in the prosecutorial mindset condition ($M = 2.11$) as compared to the generic condition ($M = 2.80$), $t(155) = 2.52$, $p = 0.012$, $d = 0.40$. This pattern of responses is consistent with blame responses: Liberals blame less in the prosecutorial mindset and this is reflected in their decreased perceptions of freedom of action. Liberals' perceptions of freedom of action in the intentional or implantation conditions were unaffected by the prosecutorial mindset manipulation (ts < 1; ps > 0.641).

As for conservatives, pairwise comparisons revealed that, in the biological defect condition, perceptions of freedom of action are significantly higher in the prosecutorial mindset ($M = 3.50$) versus the generic condition ($M = 2.59$), $t(155) = 3.09$, $p = 0.002$, $d = 0.49$. As with liberals, this pattern parallels their blame judgments. Furthermore, for conservatives, in the implantation condition, perceptions of freedom of action were significantly higher in the generic condition ($M = 4.55$) versus the prosecutorial mindset manipulation ($M = 4.02$), $t(155) = 2.16$, $p = 0.032$, $d = 0.34$. Interestingly, this result has no parallel in conservatives' blame responses. Conservatives' perceptions of freedom of action in the intentional condition were unaffected by the prosecutorial mindset manipulation ($t < 1$; $p > 0.460$).

Next, we analyzed perceived control of self-formation. There was a significant main effect of transgression narrative, $F(2, 155) = 6.01$, $p = 0.003$. Pairwise comparisons revealed that perceptions of control of self-formation were significantly higher in the intentional condition ($M = 3.64$) compared to the implantation ($M = 2.63$), $t(155) = 6.03$, $p < 0.001$, $d = 0.96$, and biological defect conditions ($M = 2.63$), $t(155) = 5.94$, $p < 0.001$, $d = 0.95$. Perceptions were not significantly different in the implantation condition compared to the biological defect condition ($t < 1$; $p > 0.983$). There was also a significant main effect of ideology, $F(1, 155) = 5.02$, $p = 0.026$. Again, there was a three-way interaction of transgression narrative, prosecutorial mindset manipulation, and ideology for perceptions of control over self-formation, $F(2, 155) = 4.39$, $p = 0.014$ (See Figures 12 and 13 for pattern of means presented separately for liberals and conservatives). For liberals, perceptions of control of self-formation were unaffected by the prosecutorial mindset manipulation in each of the transgression narrative conditions (ts < 1; ps > 0.396). Therefore, for liberals in the implantation condition, perceptions of control of self-formation did not parallel blame responses, suggesting a "direct effect" of motivation on blame responses.

For conservatives, there was no effect of the prosecutorial mindset manipulation on perceptions of control over self-formation in the intentional or implantation conditions (ts < 1; ps > 0.509). In the biological defect condition, however, perceptions were significantly higher in the prosecutorial mindset condition ($M = 3.59$) versus the generic condition ($M = 2.28$), $t(155) = 3.51$, $p = 0.001$, $d = 0.56$. Therefore, for conservatives in the biological defect condition, it appears that differences in blame responses as a result of the motivation manipulation resulted in analogous changes in ascriptions of control of

self-formation, presumably to justify the increased blame responses. In sum, conservatives blame James more in the biological defect condition when in the prosecutorial mindset, and this is paralleled by changes in their perceptions of James' freedom of action and control of his self-formation.

Next, we used path analysis to test possible mediational pathways. We will focus on the biological defect condition, as that is the only condition in which meditation of blame judgments by free will perceptions is possible (e.g., in the implantation condition, liberals blamed differentially based on mindset, but their free will perceptions showed no such pattern and thus cannot be mediating their blame judgments; in the implantation condition, conservatives had different free will perceptions based on mindset, but their blame judgments showed no such pattern, and thus there is no blame effect to be meditated).

We tested the extent to which the effect of the prosecutorial mindset on blame in the biological defect condition was mediated by perceptions of freedom of action and, in a separate analysis, by perceptions of control of self-formation. Because our ANOVA results indicated that the influence of mindset on blame and free will perceptions differed as a function of ideology, we conducted a moderated mediation analysis. We computed the analysis using the macros available at (http://www.afhayes.com/spss-sas-and-mplus-macros-and-code.html; Preacher, Rucker & Hayes, 2007) and we tested their Model 2. Please see Figure 14 for the models we tested. These models stipulated that the mindset manipulation affected free will perceptions (freedom of action, control of self-formation), that the effect of mindset on free will perceptions was moderated by ideology, and that free will perceptions contributed to blame.

In the model examining perceived freedom of action, we uncovered a significant interaction between mindset and ideology for predicting freedom of action, $t(49) = 2.74$, $p = .008$. As can be seen in Figure 15 (and consistent with the ANOVA results above), the mindset manipulation had quite different effects on perceived freedom of action among liberals versus conservatives. Among conservatives, perceived freedom of action was higher in the prosecutorial mindset condition than in the generic condition, whereas among liberals this pattern was reversed. As can be seen underneath the path models in Figure 15, the mediated effect of mindset on blame traveling through freedom of action was significant for both conservatives and liberals according to bootstrap methods (and marginal according to a Sobel test). Thus, when judging a harm-doer with a biological defect, being in a prosecutorial mindset heightens blame among conservatives by increasing their judgment that the harm-doer has freedom of action, but decreases blame among liberals by lowering their judgment that the harm-doer has freedom of action.

Next, we examined control of self-formation as a mediator. There was a significant interaction between mindset and ideology for predicting control of self-formation, $t(49) = 2.62$, $p = 0.011$. As can be seen in Figure 16 (and consistent with the ANOVA results above), the mindset manipulation influenced perceptions of control of self-formation differently among liberals versus conservatives. Among conservatives, perceived control of self-formation was higher for those in the prosecutorial mindset than in the generic condition, whereas among liberals perceived control of self-formation was unaffected by the prosecutorial mindset manipulation. As can be seen under the path models presented in Figure 16, the mediated effect of mindset on blame, traveling

through control of self-formation perceptions, was significant for conservatives—but not for liberals—according to bootstrap and a Sobel test.

Therefore, free will perceptions significantly mediate blame responses in the biological defect condition. Conservatives show significant mediated pathways for both freedom of action and control of self-formation but liberals only show a significant mediated pathway for freedom of action perceptions. In contrast to these meditational pathways, results in the implantation condition suggest more of a "direct effect" of the prosecutorial mindset manipulation on liberals blame responses (i.e., no mediation by free will perceptions).

Discussion

In Experiment 3, we crossed our transgression narrative manipulation with a prosecutorial mindset manipulation. Prior to reading about James, participants read about the company James works for: Either generic information about how the company has changed over time or a prosecutorial mindset manipulation, which emphasizes how harassment and bullying at the company has skyrocketed. Results suggest that the mindset manipulation was influential, but in different ways for liberals and conservatives (in line with existing literature by Tetlock and colleagues, 2007). In the implantation condition, liberals fit our predictions: They blamed more when in the prosecutorial rather than generic mindset. Interestingly, however, this increase in blame responses was not associated with any changes in free will perceptions, suggesting a "direct effect" of the motivation manipulation (or mediation by an unmeasured variable). In contrast, in the biological defect condition, liberals blamed less in the prosecutorial mindset. We think this is because, in the biological defect condition, liberals may think that not only does

James have a brain tumor, but also he works in an office culture that facilitates inappropriate behavior—making it very difficult for a biologically impaired person such as James to behave more positively. This decreased blaming by liberals was mediated by reduced perceptions of James' freedom of action.

Conservatives showed a different pattern of responses. They were unaffected by the prosecutorial mindset manipulation in the intentional and implantation conditions. In the biological defect condition, however, conservatives blamed more in the prosecutorial mindset than in the generic condition. This increased blaming was mediated by both increased freedom of action and increased control of self-formation perceptions. While this finding was somewhat unexpected [after all, when James has a brain tumor he is consistently blamed less, and Tetlock et al. (2007) showed that blaming of those with biological defects was unaffected by a prosecutorial mindset manipulation], it is still, in a sense, in line with our thinking. Specifically, the brain tumor information is a type of deterministic scenario. And, perhaps conservatives find the deterministic implications less than fully compelling (e.g., "*James could use willpower to overcome the effects of his brain tumor*"), creating malleability in their judgments of that type of deterministic scenario. Why conservatives would find an unfortunate history to be more consistently compelling than a brain tumor in terms of blame reduction is a mystery.

We found that the prosecutorial mindset manipulation was effective for liberals in the implantation condition but not conservatives. We think this divergence in responses may have to do with the nature of our vignettes. Tetlock and colleagues (2007) have found that the liberals are more susceptible to the prosecutorial mindset manipulation when it pertains to punishing people in positions of power. Therefore, since James is the

vice president of the company he works for, he is likely seen as largely contributing to the harassment problems at the company in the prosecutorial mindset—and liberals are particularly influenced by this kind of information. In contrast, our prosecutorial mindset manipulation may have been more effective in the implantation condition if the scenario pertained to street crime, which conservatives are more likely to be swayed by (Tetlock et al., 2007). Future replications of this study should be aimed at confirming this hypothesis: Does James' position in the company hierarchy influence the way he is perceived by different groups? Furthermore, is it this perception that amplifies blame responses for liberals in the implantation condition?

Another interesting question is why some deterministic scenarios are more compelling to some groups over others. Why are conservatives more willing to blame James in the biological defect condition but not the implantation condition, when in the prosecutorial mindset? It may be the case that the specific nature of having a brain tumor or neurological defect is considered a less 'legitimate' excuse for poor behavior for conservatives—but why? Similarly, liberals' intuitions about determinism and moral responsibility were more malleable than those of conservatives in the implantation condition. Is a person's unfortunate personal history less compelling reason for liberals as to why a person should not be considered morally responsible? This thought process does not seem likely. Additional studies are needed to further understand when a deterministic scenario will be compelling and legitimate for some groups or individuals but not for others.

General Discussion

The philosophy literature highlights how perceptions of free will and determinism are influential in determining ascriptions of moral responsibility. However, while much current research attempts to determine whether people are compatibilist or incompatibilist, we offered the prediction that people can flip-flop between these two positions based on their intuition that free will resides along a continuum. That is, we have argued that people are able to maintain malleable intuitions regarding whether deterministic forces diminish moral responsibility. Our previous work (Gill & Cerce, 2013) has suggested that deterministic information regarding a wrongdoer's unfortunate history lowers blame of the wrongdoer, mediated by the sense that he or she lacked control over self-formation. In the present studies, our focus was on the possibility that people would be inconsistent in terms of using information about an unfortunate history in this way. We predicted that sometimes, in contrast to the finding of Gill and Cerce (2013), information about an unfortunate history would not much have much of an effect on moral responsibility judgments.

We developed predictions about two processes that would moderate the tendency for historical information to reduce blame. One process involves activating the idea that the wrongdoer could be trying harder—utilizing his or her capacity for freedom of action—to behave more appropriately. We reasoned that activation of this idea should have no effect in the biological defect condition because people view biological defects as taking away freedom of action (Gill & Cerce, 2013), and thus the wrongdoer has no such capacity to utilize. We also expected that activation of this idea should have no effect in the intentional condition because blame is already "at ceiling" there. We

expected activation of this idea to selectively increase blame in the implantation condition because it serves as a reminder that, although the wrongdoer never set out to become a bad person, he or she is fully capable in the present of changing course and can be considered blameworthy if he or she fails to properly use that capacity. We also predicted that placing people in a prosecutorial mindset would weaken the tendency for blame to be reduced by information about an unfortunate history. The rationale for such a prediction is that the prosecutorial mindset shifts a perceiver's focus from the attributes of the wrongdoer to the need to restore social order by blaming and punishing bad behavior. With this shift in focus, wrongdoer attributes that would normally reduce blame (e.g., lack of control over self-formation) receive less weight.

We presented three experiments that tested these ideas. All three studies varied the transgression narrative associated with James, an office bully. In one condition, James bullied based on a biological defect. In a second condition, James bullied based on morally offensive intentions. In a third condition, James bullied based on these same morally offensive intentions but the intentions were implanted in him by an unfortunate personal history. Our key prediction across the three experiments was that manipulations derived from the reasoning in the preceding paragraph—crossed with our transgression narrative manipulation—would selectively moderate blame and moral responsibility judgments in the implantation condition (See Figure 1). We anticipated less influence of the manipulations within the biological defect and intentional conditions.

In Study 1, we crossed our transgression narratives with a manipulation highlighting James' refusal to change. Our goal with the refusal to change manipulation was to selectively activate freedom of action perceptions, which would result in increased

ascriptions of blame and moral responsibility in the implantation condition. We did not find these effects. Furthermore, results suggest that our refusal to change manipulation did not increase freedom of action perceptions as intended. Instead, it increased control of self-formation perceptions.

In Study 2, we again crossed our existing transgression narratives with a manipulation highlighting one's own ability to overcome predispositions. The manipulation was intended to increase perceptions of freedom of action and result in increased blame in the implantation condition. Blame responses did not show this predicted pattern. Additionally, the manipulation resulted in decreased perceptions of control of self-formation—not increased perceptions of freedom of action. However, we did find some evidence for our predicted effect: Participants indicate greater ascriptions of "encourage him to change" punishment responses after the overcoming predispositions task in the implantation condition. But this pattern was reversed for the biological defect and intentional conditions.

Study 3 manipulated whether participants were in the prosecutorial mindset prior to reading the transgression narratives. This mindset manipulation was meant to motivate observers to selectively attend to information relevant to culpability (to fulfill the goal of holding transgressors responsible in order to maintain order in society) and, thus, ignore deterministic information regarding an unfortunate personal history. Results suggest that some groups show this effect: Liberals, in the implantation condition blame marginally more when in the prosecutorial mindset. In contrast, conservatives, in the biological defect condition, blame significantly more in the prosecutorial mindset (and liberals blame less in the biological defect/prosecutorial mindset condition).

In sum, results across three studies are mixed with regard to whether people have inconsistent intuitions about determinism and moral responsibility. Some of our results (i.e., liberals in the prosecutorial mindset) indicate that people will ascribe moral responsibility even in an implantation (i.e., deterministic) scenario. The punishment findings in Study 2 corroborate these results: Participants indicate increased "encourage him to change" punishment responses in the implantation condition after completing the overcoming predispositions task. However, we found this pattern inconsistently across the three studies presented here.

At this juncture, we have not found conclusive evidence in opposition to our hypotheses. Specifically, although Studies 1 and 2 did not show the predicted pattern of blame responses, it is clear that our freedom of action reminder manipulations were not manipulating the intended constructs. In both studies, the intended freedom of action reminder actually manipulated control of self-formation perceptions. In Study 1, the refusal to change manipulation resulted in increased perceptions of control of self-formation (but unevenly across conditions). In contrast, in Study 2, the overcoming predispositions resulted in decreased perceptions of self-formation (again, unevenly across conditions). However, blame responses only minimally reflected these differences in free will perceptions. Despite the fact that the results of Study 1 and 2 suggest that our refusal to change and overcoming predispositions manipulations influence perceptions of control of self-formation, this free will perception did not transfer to blame responses, likely because the effect of the manipulation was, by all indicators, weak.

Furthermore, these manipulations were not selectively influencing free will perceptions and moral responsibility ascriptions only in the implantation condition. We

consistently found significant differences in free will perceptions and punishment responses across all three transgression vignettes after the refusal to change and overcoming predispositions manipulations. This finding was unexpected: We did not expect perceptions of free will or blame ascriptions to change in the biological defect condition because, in this condition it is impossible to increase free will perceptions due to the unchangeable, uncontrollable brain tumor.

We have found some evidence that people do flip-flop between compatibilist and incompatibilist judgments (i.e., Study 2: punishment responses in implantation condition; Study 3: liberals in the implantation condition; conservatives in biological defect condition). However future studies should be focused on more effectively manipulating our reminders of freedom of action capabilities. Specifically, in Studies 1 and 2, our manipulations predominantly influenced control of self-formation ascriptions. Perhaps manipulations focused on James' ability to exert willpower in other aspects of his life (e.g., forgoing a short term gain in order to achieve a long term goal), or a manipulation that highlights James' mental activities when vacillating between two behavior options (e.g., bully or not bully) would selectively activate freedom of action perceptions. Or, it may be the case that motivational manipulations, such as the prosecutorial mindset study presented here, are better at selectively focusing participants' attention away from the deterministic aspects of the scenario. Perhaps an experiment that manipulated relational mindsets (i.e., communal sharing versus authority ranking; Fiske, 1992) could result in increased blame responses in the implantation condition. In other words, it may be that the framing of the scenario [e.g., Is my goal to consider the well-being of others (communal sharing) or to compete with others (authority ranking)?] that determines

whether the intuition that deterministic information diminishes moral responsibility is relevant in this context.

Conclusion

Beliefs about determinism and moral responsibility are not black and white. Instead, sometimes knowledge of deterministic factors can diminish moral responsibility and at other times, it does not. In fact, people sometimes appear to be susceptible to motivations that can sway their ascriptions of moral responsibility—despite a deterministic scenario. However, this largely depends on preexisting characteristics of the individual (e.g., political ideology) and the nature of the scenario (e.g., in some cases unfortunate personal history no longer diminishes blame, in other cases a biological defect no longer diminishes blame). We see our work as taking an important theoretical step in terms of arguing that laypersons maintain flexible intuitions regarding determinism and moral responsibility. However, based on our findings, their intuitions are inconsistently activated based on the information available. Future experiments are necessary to solidify whether there is an underlying consistency and whether it is more strongly moderated by reminders of freedom of action capacities or motivational factors. We have found some evidence to support the claim that laypersons may appear compatibilist in some scenarios but incompatibilist in others when they are, in fact, neither. However, the extent to which they consistently apply compatibilist and incompatibilist responses remains an enigma.

References

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological bulletin*, *126*(4), 556–74.

Feltz, A., Cokely, E. T., & Nadelhoffer, T. (2009). Natural compatibilism versus natural incompatibilism: Back to the drawing board. *Mind & Language, 24*(1), 1-23.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review, 99*(4), 689.

Gill, M. J., Andreychik, M. R., & Getty, P. D. (2013). More than a lack of control external explanations can evoke compassion for outgroups by increasing perceptions of suffering (independent of perceived control). *Personality and Social Psychology Bulletin, 39*(1), 73-87.

Gill, M., & Cerce, S. (2013; in preparation) He never willed to will that: Historical information about an immoral actor attenuates blame by altering particular free will perceptions (control of self-formation), while leaving others intact (freedom of action).

Haidt, J. (2001). The emotional dog and its rational tail: a social intuitionist approach to moral judgment. *Psychological review, 108*(4), 814.

Hume, D. (1738). *A treatise of human nature*. London: John Noon.

Kant, I. (1781). *Critique of pure reason*. G, Bell.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754-770.

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2011). Moral, cognitive, and social: The nature of blame. *Social thinking and interpersonal behavior*, 313-331.

Malle, B. F., & Knobe, J. (1997). The folk concept of intentionality. *Journal of Experimental Social Psychology*, *33*(2), 101-121.

Nahmias, E., Coates, D. J., & Kvaran, T. (2007). Free will, moral responsibility, and mechanism: Experiments on folk intuitions. *Midwest Studies in Philosophy, 31*(1), 214-242.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner 1, J. (2005). Surveying freedom: Folk intuitions about free will and moral responsibility. *Philosophical Psychology, 18*(5), 561-584.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*(4), 663-685.

Pizarro, D. A., & Tannenbaum, D. (2011). Bringing character back: How the motivation to evaluate character influences judgments of moral blame. *The social psychology of morality: Exploring the causes of good and evil*, 91-108.

Pizarro, D., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: the role of perceived metadesires. *Psychological science*, *14*(3), 267–72. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12741752

Preacher, K. J., Rucker, D. D., & Hayes, A. F. (2007). Assessing moderated mediation hypotheses: Theory, methods, and prescriptions. *Multivariate Behavioral Research, 42*, 185-227.

Strawson, P. (1963). Freedom and resentment. *Perspectives on Moral Responsibility*, 67-100.

Tetlock, P. E., Visser, P. S., Singh, R., Polifroni, M., Scott, A., Elson, S. B., & Rescober, P. (2007). People as intuitive prosecutors: The impact of social-control goals on

attributions of responsibility. *Journal of Experimental Social Psychology*, *43*(2), 195-209.

Watson, G. (1987). Free action and free will. *Mind*, *96*(382), 145-172.

Weiner, Bernard. (2006). A theory of social motivation and justice: Logic and development. *Social Motivation, Justice, And The Moral Emotions: An Attributional Approach* (pp. 1–42).

Weiner, Bernard, & Kukla, A. (1970). An attributional analysis of achievement. *Journal of Personality and Social Psychology*, *15*(1), 1–20.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: studies in the attribution of moral responsibility. *Cognition*, *100*(2), 283–301. doi:10.1016/j.cognition.2005.05.002

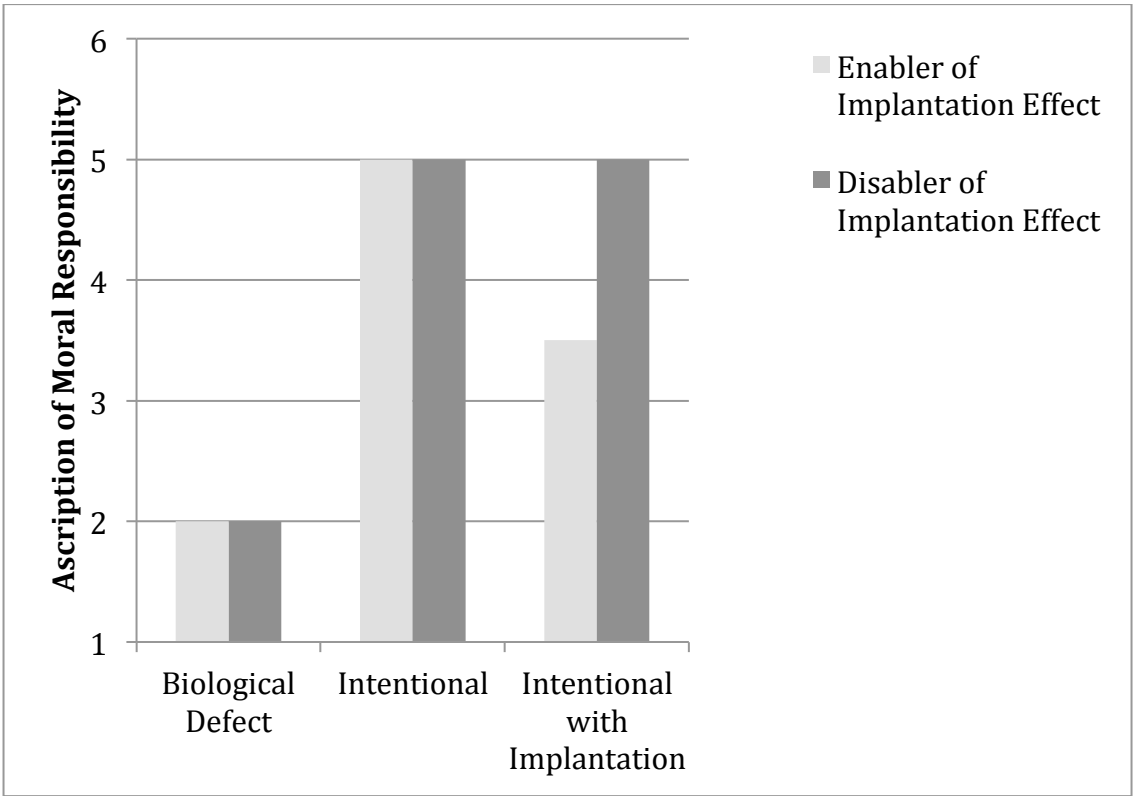


Figure 1. Idealized predictions for Experiments 1-3. "Enabler of Implantation Effect" bars represent findings from Gill & Cerce (2013): Blame responses were significantly higher in the intentional condition compared to the implantation condition (despite equivalent intentionality perceptions). "Disabler of Implantation Effect" bars represent predicted blame responses of the studies presented here: When presented with a reminder of freedom of action or when experiencing heightened blame motivation, blame responses in the implantation condition will be comparable to the intentional condition—thus disabling the implantation effect.

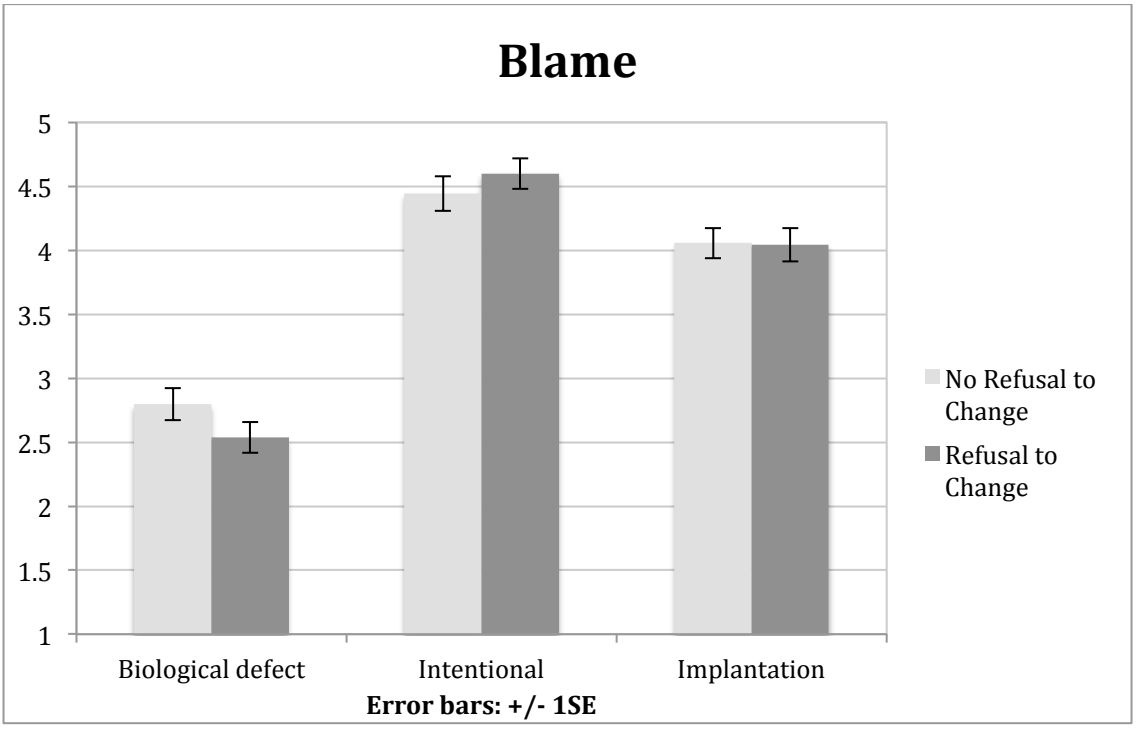


Figure 2. Blame responses as a function of transgression narrative and refusal to change manipulation in Experiment 1.

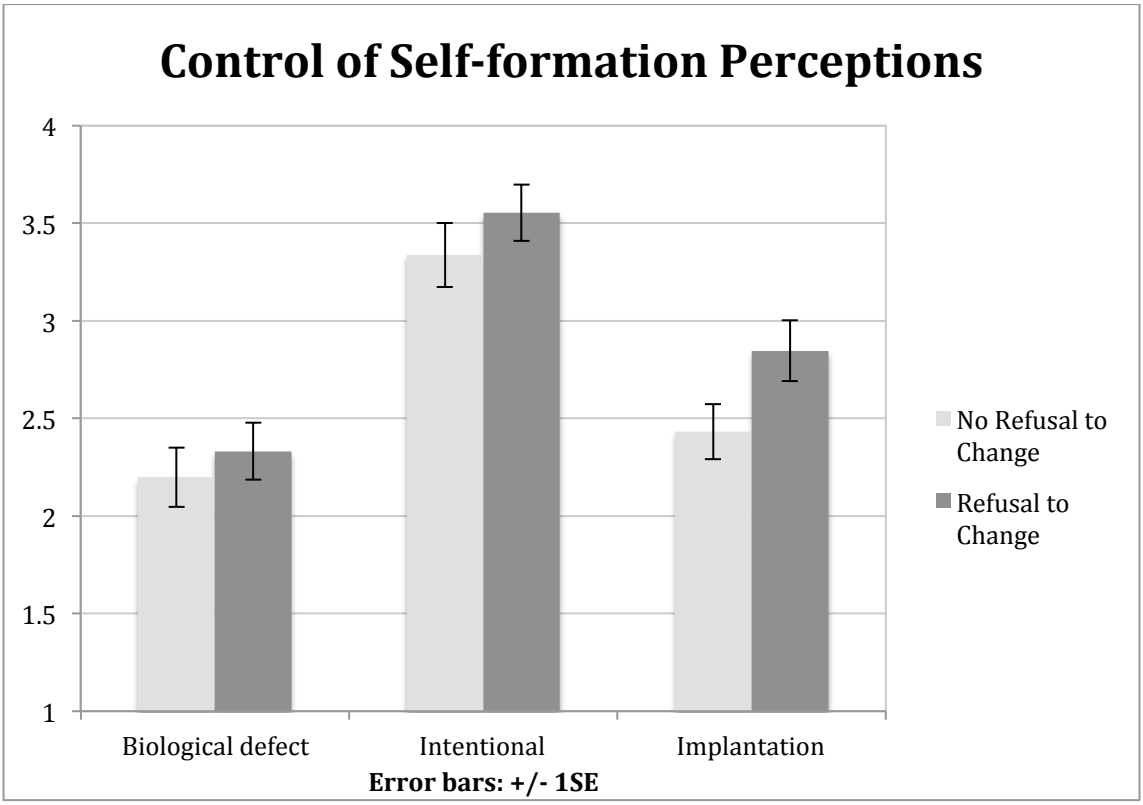


Figure 3. Control of self-formation perceptions as a function of transgression narrative and refusal to change manipulation in Experiment 1.

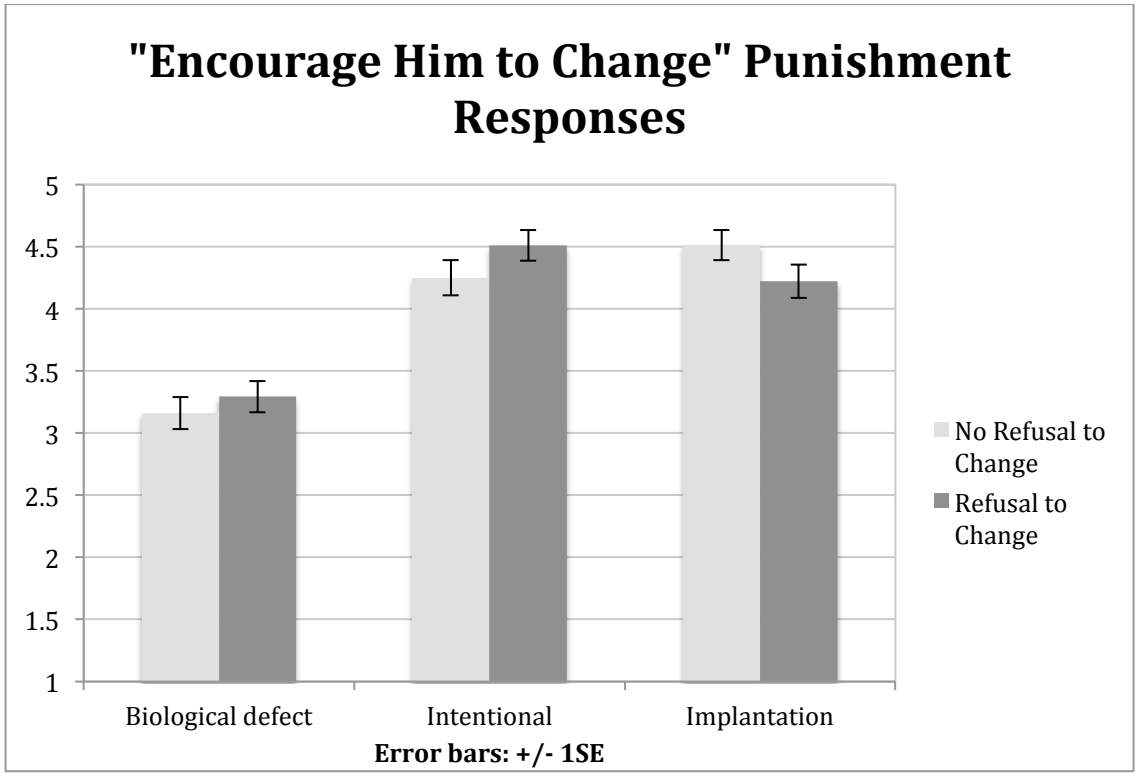


Figure 4. "Encourage him to change" punishment responses as a function of transgression narrative and refusal to change manipulation in Experiment 1.

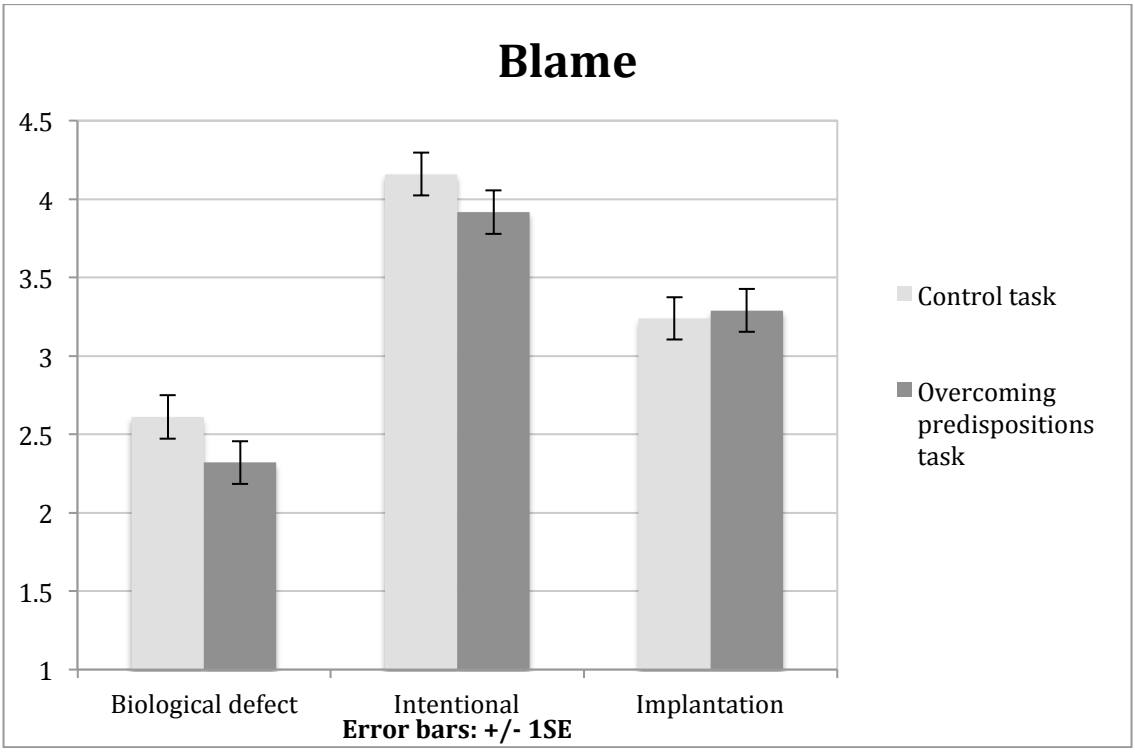


Figure 5. Blame responses as a function of transgression narrative and overcoming predispositions manipulation in Experiment 2.

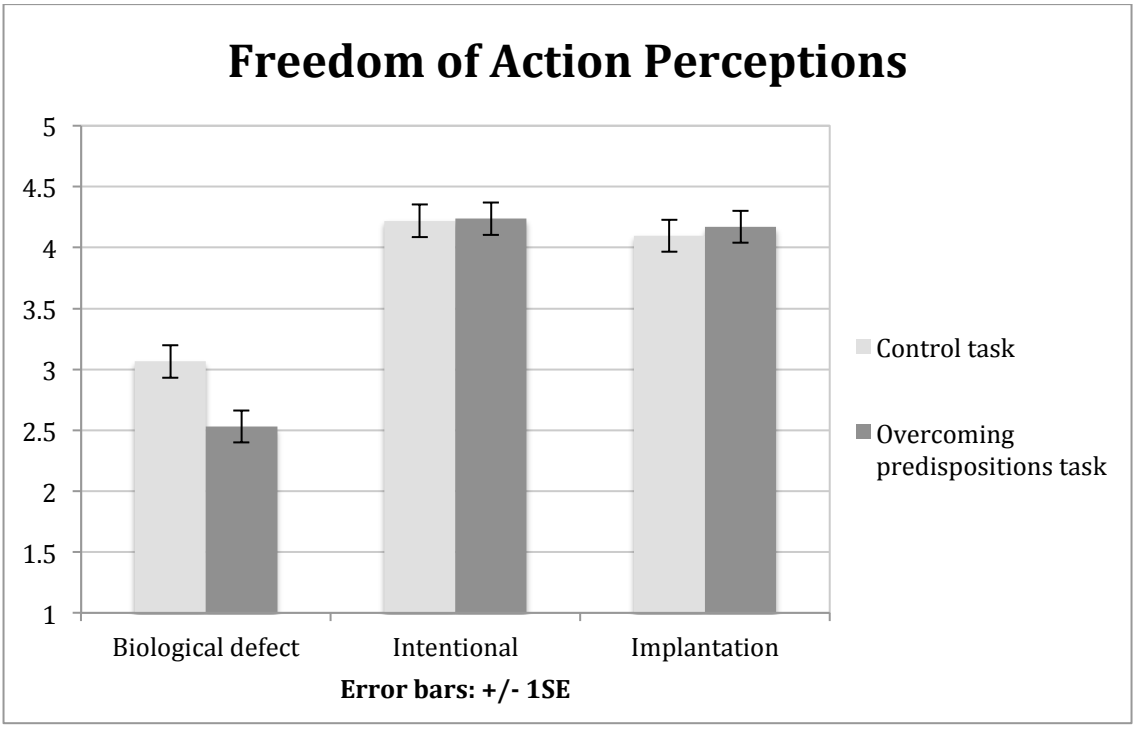


Figure 6. Freedom of action perceptions as a function of transgression narrative and overcoming predispositions manipulation in Experiment 2.

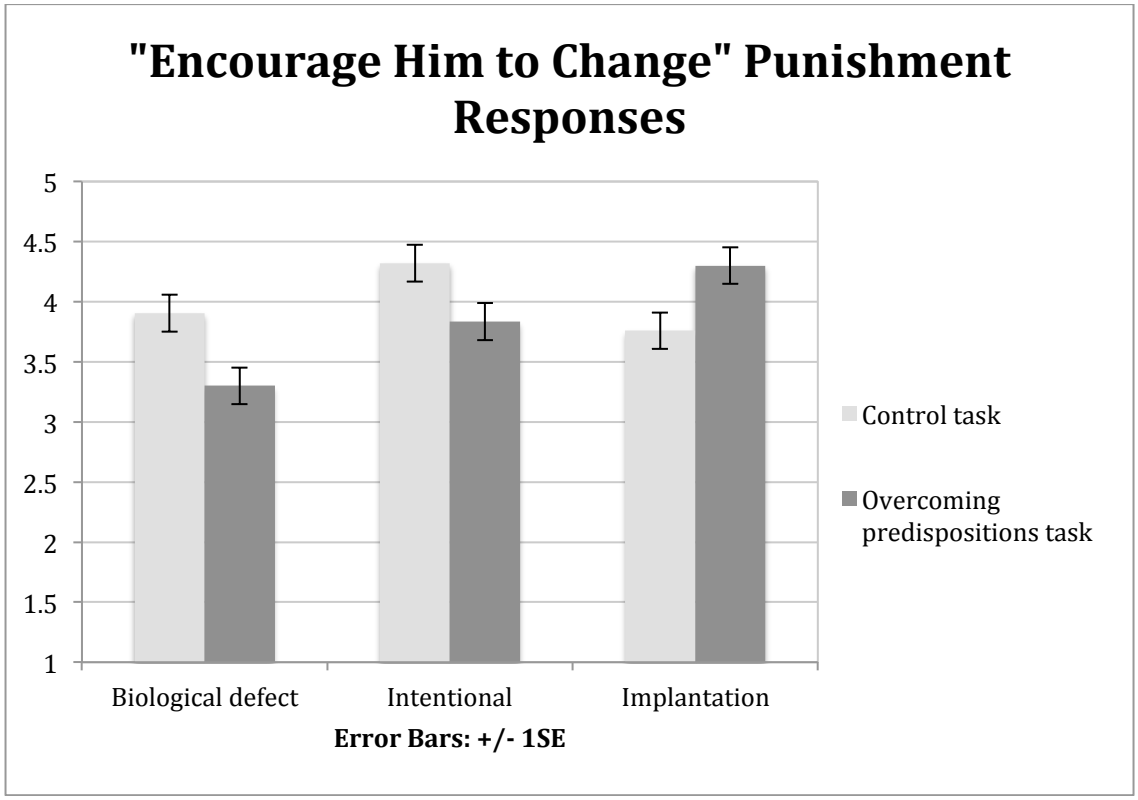


Figure 7. "Encourage him to change" punishment responses as a function of transgression narrative and overcoming predispositions manipulation in Experiment 2.

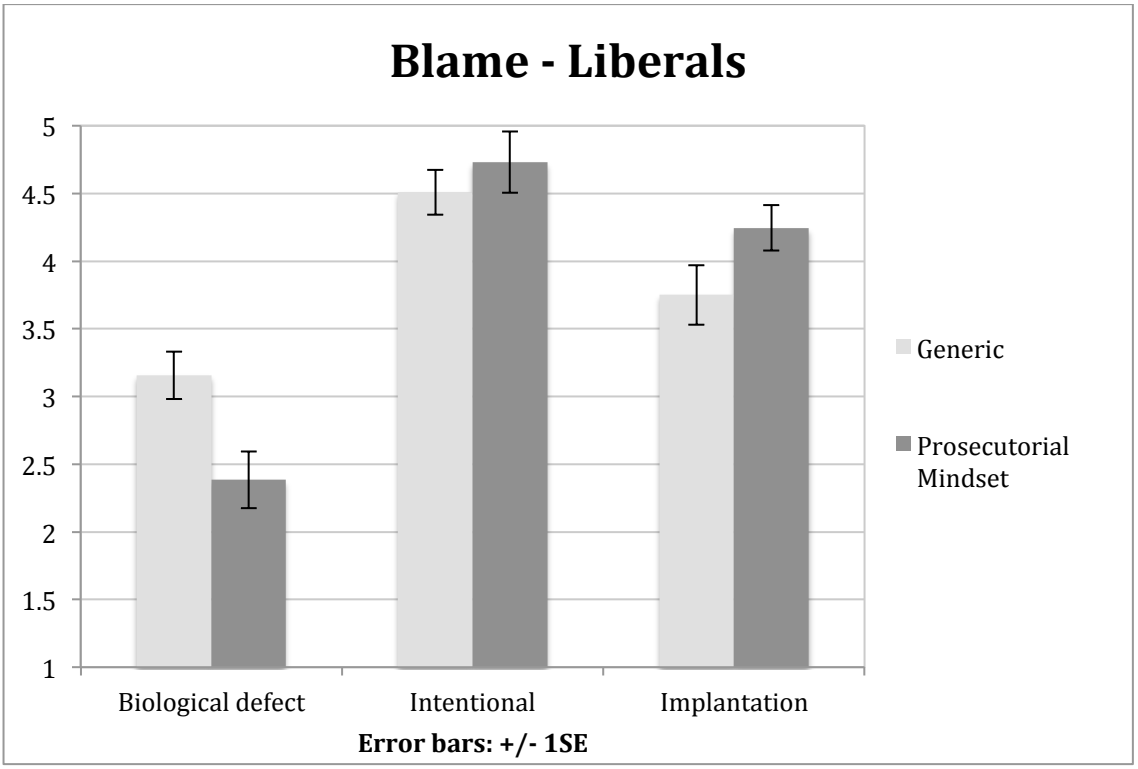


Figure 8. Blame responses for liberals as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

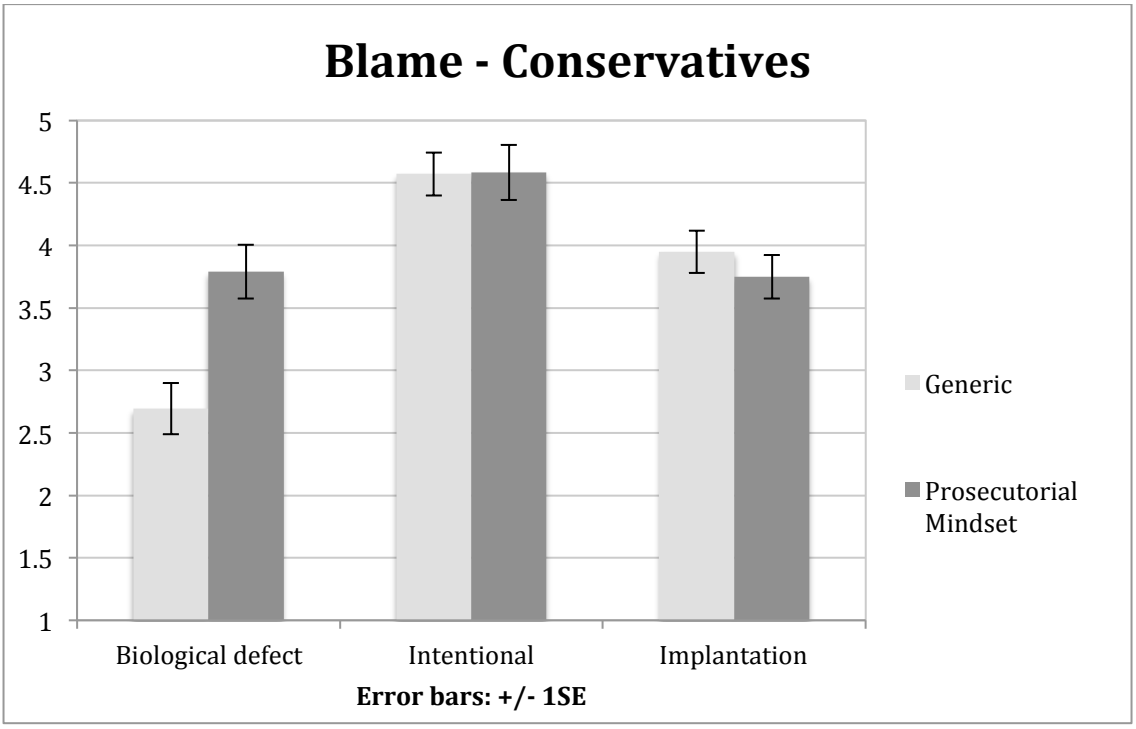


Figure 9. Blame responses for conservatives as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

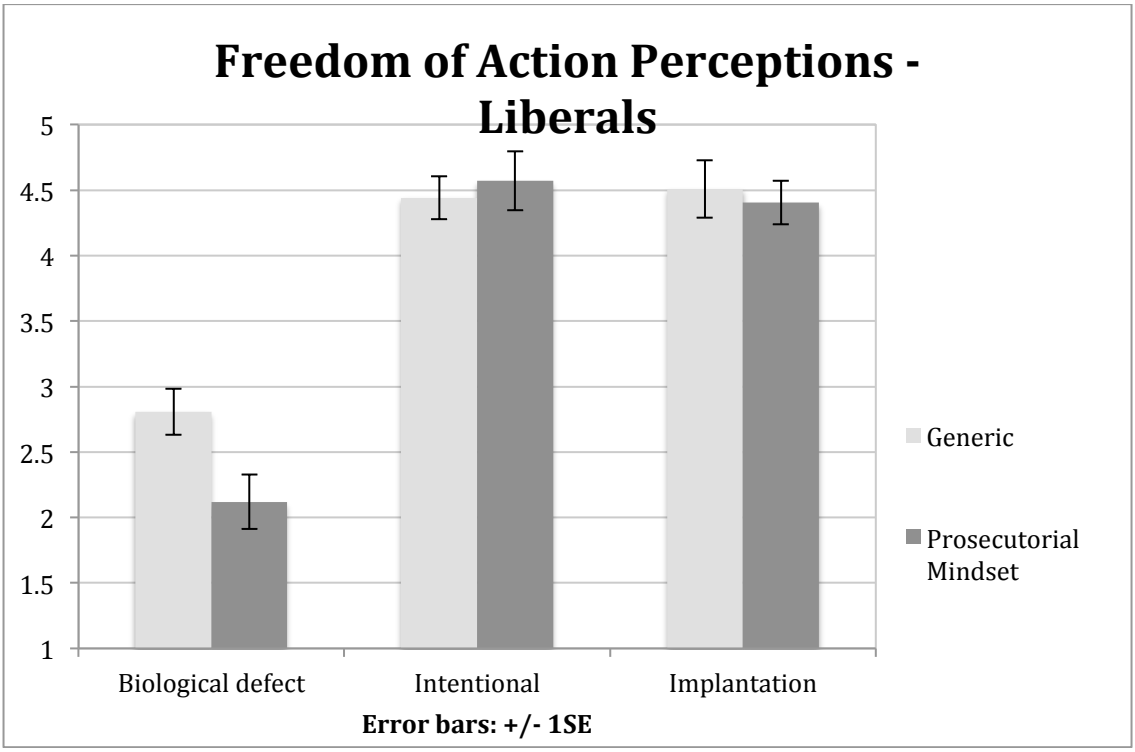


Figure 10. Freedom of action perceptions for liberals as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

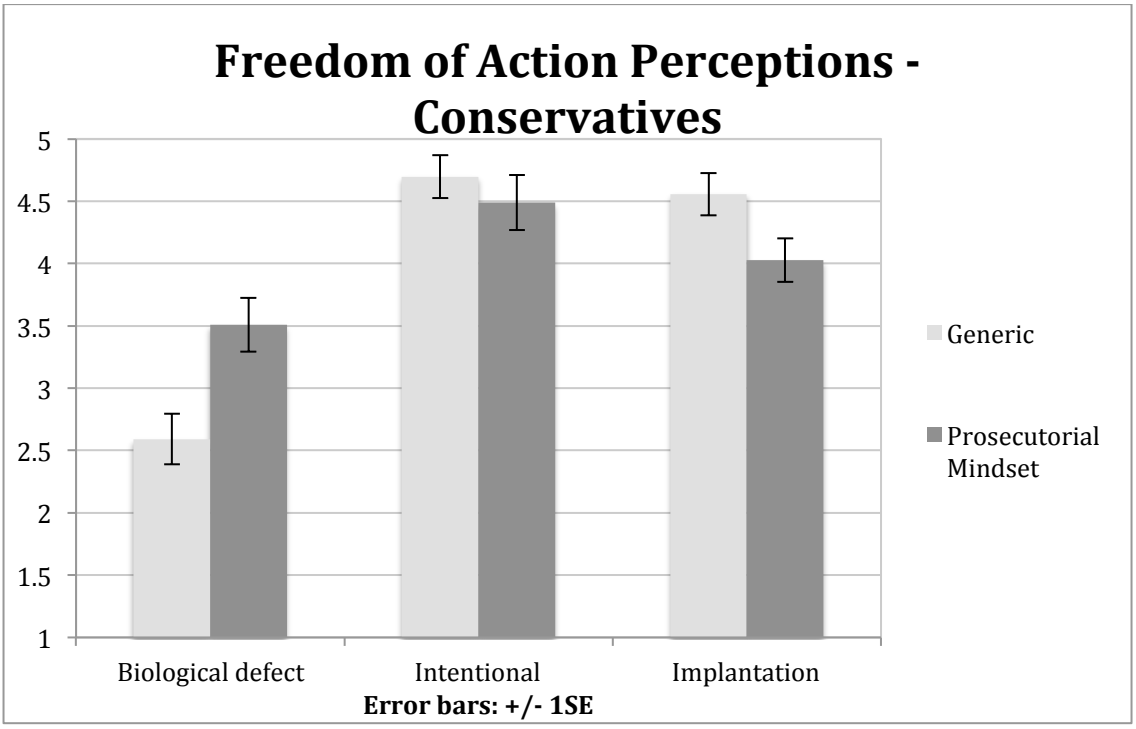


Figure 11. Freedom of action perceptions for conservatives as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

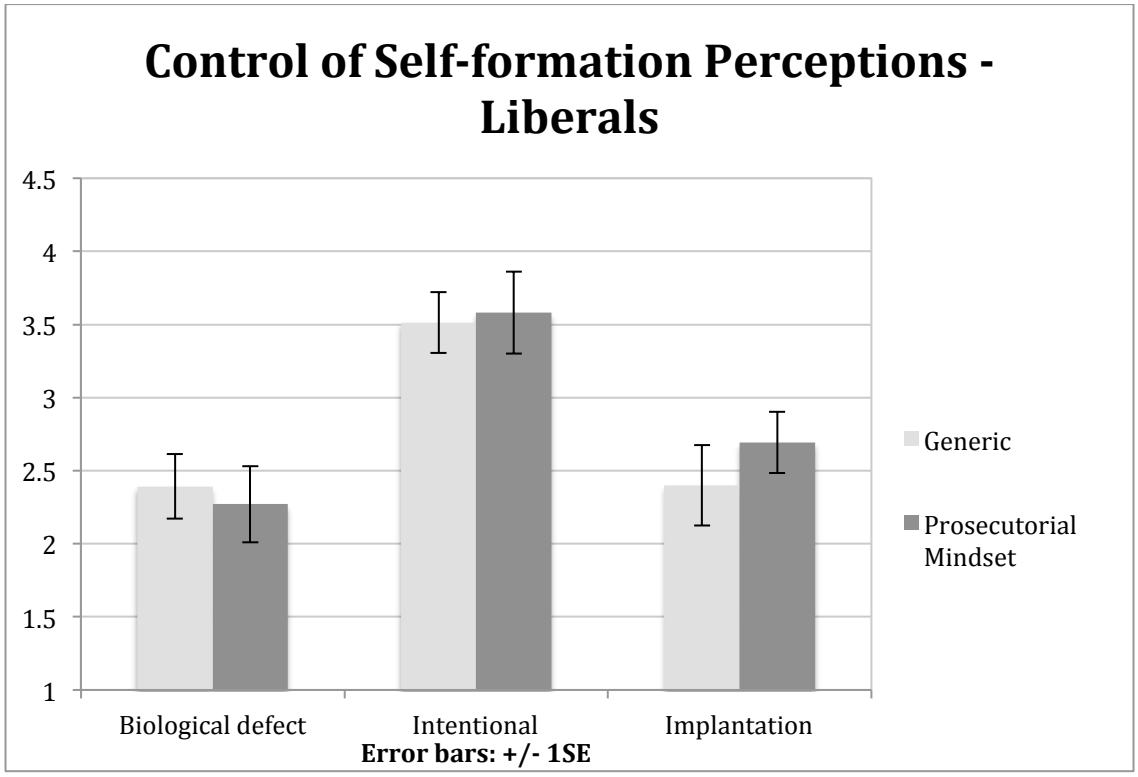


Figure 12. Control of self-formation perceptions for liberals as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

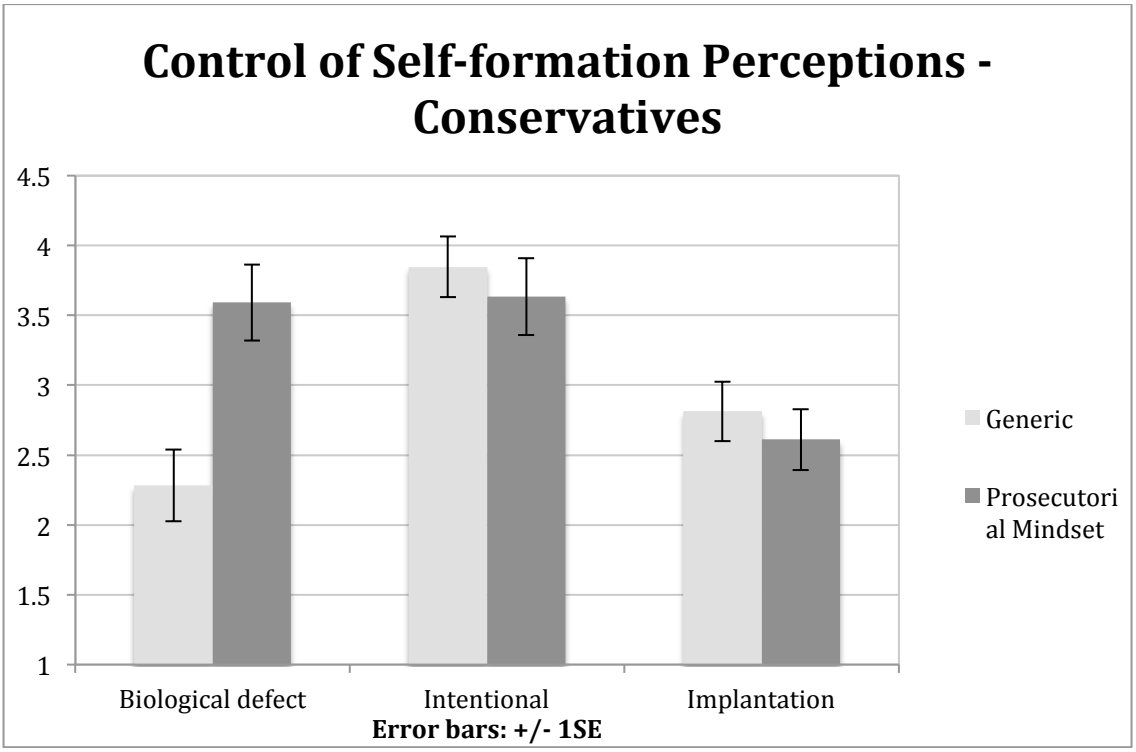


Figure 13. Control of self-formation perceptions for conservatives as a function of transgression narrative and prosecutorial mindset manipulation in Experiment 3.

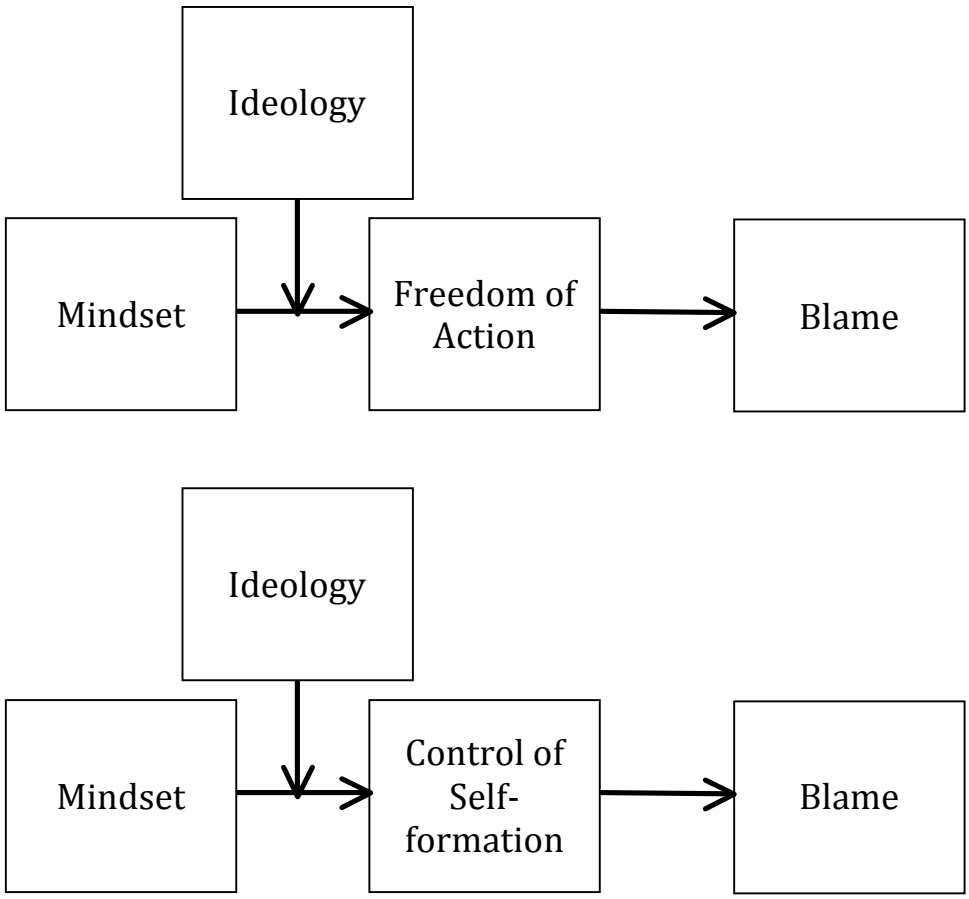
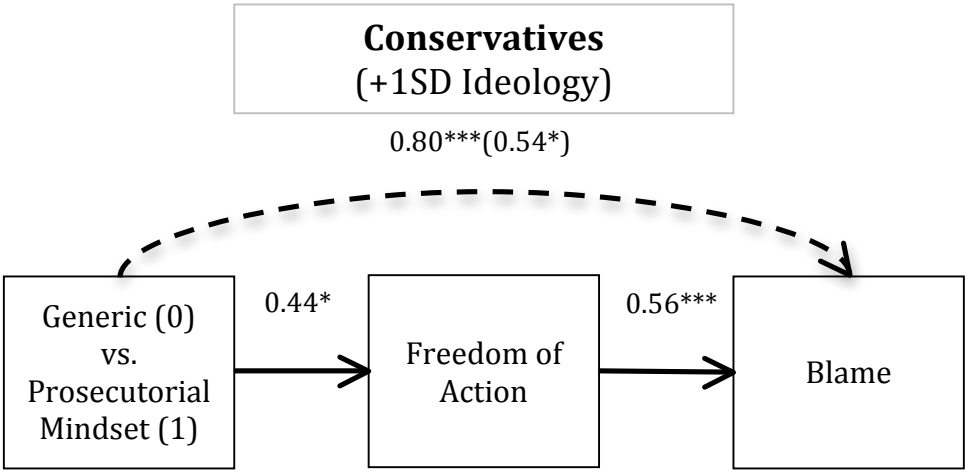
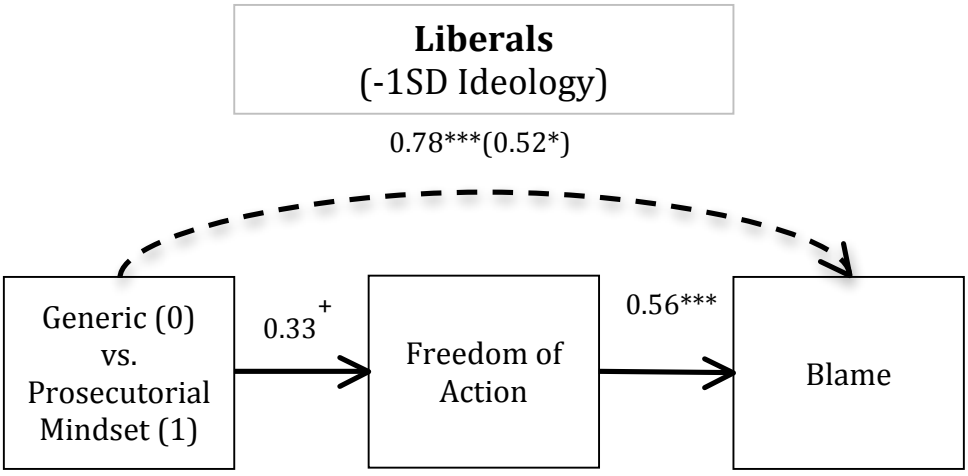


Figure 14. Moderated mediation models tested in Experiment 3. These models prescribe that the prosecutorial mindset manipulation and ideology interact to predict free will perceptions (freedom of action and control of self-formation) and that free will perceptions predict blame.

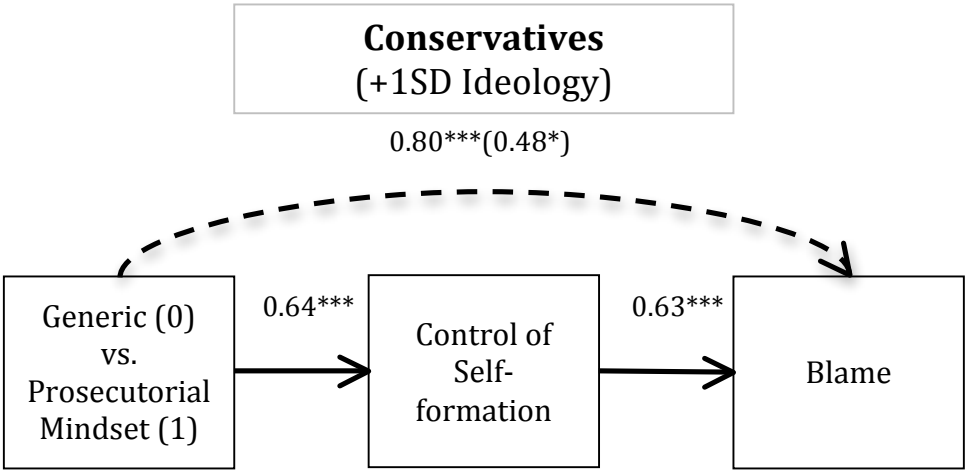


*p < 0.05; ** p < 0.01; *** p < 0.001
Sobel z = 1.80, p = 0.07
Bias Corrected Bootstrap 95% CI (0.06, 0.86)

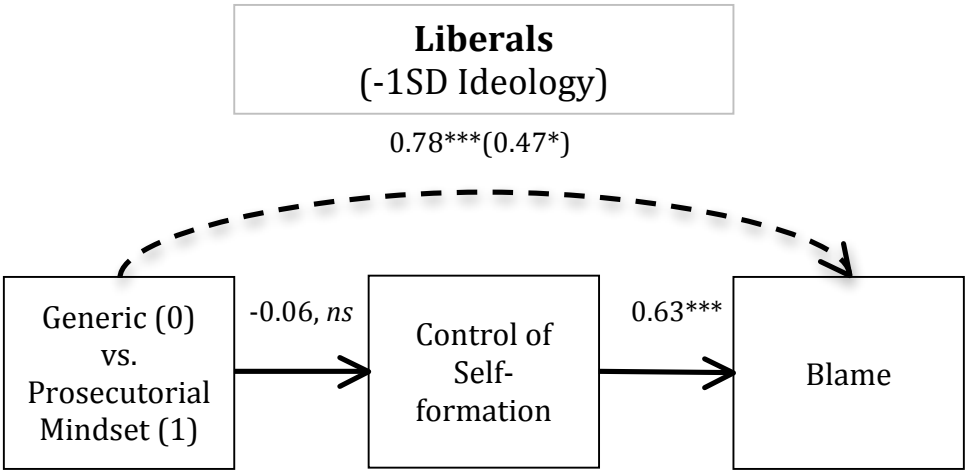


+ p < 0.1 *p < 0.05; ** p < 0.02; *** p < 0.001
Sobel z = -1.54, p = 0.12
Bias Corrected Bootstrap 95% CI (-0.72, -0.001)

Figure 15. The mediated effect of mindset manipulation on blame via perceived freedom of action for conservatives and liberals in Experiment 3.



*p < 0.05; ** p < 0.02; *** p < 0.001
Sobel z = 2.69, p = 0.007
Bias Corrected Bootstrap 95% CI (0.25, 1.26)



*p < 0.05; ** p < 0.02; *** p < 0.001
Sobel z = -0.33, p = 0.74
Bias Corrected Bootstrap 95% CI (-0.46, 0.27)

Figure 16. The mediated effect of mindset manipulation on blame via perceived control of self-formation for conservatives and liberals in Experiment 3.

Appendix A

Intentional Transgression narrative

James is the Vice President of a moderately sized company who will often humiliate and harass members of his staff. He says unkind things and implies that employees are unintelligent or incompetent. Employees avoid James as much as possible so that they will not be his next target. Unfortunately, there are few avenues for change in the company as James is the owner's son and employees are hesitant to file complaints against him.

For James, bullying is a choice. His choice is driven mainly by a desire to be superior to others. He thinks that putting others down is "good sport" and that it is funny when other people are embarrassed. He often plans his insults in advance, timing them to achieve maximum impact. On occasions when others have asked James to reflect on what he is doing, this has produced no positive effect. James knows that he is making others suffer, but he feels that it is legitimate to do whatever it takes to "be on top."

Implantation Transgression narrative

James is the Vice President of a moderately sized company who will often humiliate and harass members of his staff. He says unkind things and implies that employees are unintelligent or incompetent. Employees avoid James as much as possible so that they will not be his next target. Unfortunately, there are few avenues for change in the company as James is the owner's son and employees are hesitant to file complaints against him.

For James, bullying is a choice. His choice is driven mainly by a desire to be superior to others. He thinks that putting others down is "good sport" and that it is funny when other people are embarrassed. He often plans his insults in advance, timing them to achieve maximum impact. On occasions when others have asked James to reflect on what he is doing, this has produced no positive effect. James knows that he is making others suffer, but he feels that it is legitimate to do whatever it takes to "be on top."

None of the employees are aware of the abusive personal relationship between James and his father. Throughout his childhood, James' father physically and verbally abused him. James was often told that he was worthless and an idiot. Whenever James upset his father, he would be whipped with a belt or even punched. To James' young mind, the lesson was that bullying is an appropriate strategy for getting people to do what you want them to do. James came to believe that one has two options in life: "Bully" or "victim."

Biological defect Transgression narrative

James is the Vice President of a moderately sized company who will often humiliate and harass members of his staff. He says unkind things and implies that employees are unintelligent or incompetent. Employees avoid James as much as possible so that they will not be his next target. Unfortunately, there are few avenues for change in the company as James is the owner's son and employees are hesitant to file complaints against him.

A couple of years ago, James was diagnosed with a benign (i.e., *not* deadly) brain tumor that affects his frontal lobes. His doctor explained that the tumor is definitely the

cause of James' offensive social behavior. The doctor explained that tumors in that location make people aggressive and mean, even without provocation. Also, the doctor explained that the tumor would make James unable to inhibit his aggression because the self-control centers of his brain are affected. Thus, according to the doctor, it is completely impossible for James to control his unkindness.

Appendix B

Refusal to Change Manipulation

On numerous occasions, James' coworkers--and even some of his friends--have made appeals to him to change his behavior. But, James never changes. Some have become angry and tried assertive methods: *"Knock it off, a**hole!!!"* Many more have tried kindness: *"James, life is difficult enough without us being unkind to each other. Would you please think about what you are doing, think about other people's feelings, and do your best to mellow out? Everyone here would be happier if we could all be friends, or at least leave each other in peace. Wouldn't that be better?"* Despite repeated and varied persuasion attempts over the past year, James has not changed at all.

Appendix C

Blame

- I blame James for his bullying.
- I feel moral anger or disapproval toward James.
- I have strong feelings of dislike for James.
- I feel indignation toward James because of how he treats others.

Freedom of Action

- By using his human capacity for free will, James could choose to STOP being a bully.
- Although James may have a strong inclination to treat others poorly, he can use his human capacity for free will to change his behavior.
- It is possible for James to use his free will to overcome his negative impulses and behave more appropriately.
- Even if James has a strong desire to treat others poorly, he has the power to overcome that desire and treat others with respect.
- James is not at the mercy of his impulses: He can make a choice to exercise self-control of his behavior.
- James is not doomed to be a bully his whole life: By using his power of free will he can change his ways.
- Whereas James may be tempted to be mean, he does not have to yield to that temptation.

Control of Self-Formation

- James had free will in terms of initially BECOMING a bully.
- Throughout his life, James was always in control of his personality development.
- James' character traits are purely a result of him freely choosing to become who he currently is.
- Free choices that James made during his formative years are the primary reason he is the person he is today.
- James deliberately and intentionally molded himself into the type of person that would bully others.

Punishment – "Encourage him to change"

- James should receive a clear warning from a superior: You will not be eligible for a raise until you start treating others with respect.
- James should be required to see a psychologist who specializes in office bullying; the goal would be to help James see how his behavior affects others.
- James should be required to attend an office etiquette seminar so he can learn how to appropriately treat his coworkers.
- For every complaint filed against James, he should be required to meet with a human resources representative to encourage him to change his behavior.

Punishment – "Mess with him"

- It would be great to learn that one of James' coworkers to "accidentally" spilled water on his laptop, destroying its contents.
- I would be happy to hear that a project that James had been working on for over a year was "accidently" erased from his hard drive by a coworker.
- I would be pleased to hear that James' secretary was intentionally withholding important messages from him.
- It would be great if his coworkers changed the clock in his office, making him late for an important meeting.
- I wish someone would "put James in his place," and make him look foolish in front of everyone.
- It would be great if James had to pay a step price for the way he treats his staff.

Not Trying (refusal to change manipulation check) – Study 1

- James has free will—he just needs to USE it.
- James could change his ways if he would simply try harder to hear what others are telling him about his behavior.
- James stubbornly refuses to exercise the power of choice that he has over his own behavior.
- James is deliberately "thick-headed" and deliberately refuses to change.

Appendix D

Overcoming Predispositions Task

You will now complete a personal narrative task. Think of a time when you **overcame—at least for a while—a powerful bad habit** in your life. Specifically, try to think about a strong 'bad habit' or 'negative predisposition' you acquired from your family or prior life experiences. This should be a personality characteristic or behavior preference that you feel you learned over the course of your development. For example, something like 'being critical of others,' 'having a bad temper,' and 'worrying too much' are examples of traits you may have learned from your early experiences.

Once you have decided on a bad habit, briefly describe a couple of occasions when you **used your power of self-control or "will power"** to defeat or suppress the impulse to act based on the bad habit. Write about how you were able to "overpower" that bad habit through exercising your will power. Specifically, write about the choices you made in order to overcome the impulse and to avoid its influence on your actions or decisions.

Please provide as much detail as possible.

Control Task

You will now complete a personal narrative task. Think of a time when you **had free time and had to make a decision about how to fill it**. This should be a time when you had to make a decision about what to do for entertainment or otherwise you would have done nothing. For example, something like having an entire weekend off from school or work, or not making plans for spring break are examples of times you may have had free time to fill.

Briefly describe a couple of occasions when you **had to make a decision about what to do during your free time**. Specifically, write about the decision process as you were weighing alternatives in order to ultimately decide how to fill your free time.

Please provide as much detail as possible.

Appendix E

Prosecutorial Mindset

Here is the information about the company:

Thirty years ago, the Hampton & Sons Corporation was a leader in industry and manufacturing. Most Hampton employees at the time reported feeling that their company was a remarkably enjoyable place to work. Indeed, the vast majority of employees respected each other and got along well. Back then, only a tiny fraction of employees got into trouble for harassing fellow staff members, and those who did were swiftly brought to justice. The company boasted a 96% claim rate, where most complaints about harassment were quickly responded to and remediated.

Several years ago, however, the harassment rate began to soar and the rate of successful, effective responses to that harassment began to plummet. Since 2002, Hampton & Sons has been consistently ranked among the top ten most undesirable companies to work for in the United States. Last year was a record breaker: Claims regarding office bullying doubled and a record number of people (137) indicated that they suffered "severe emotional distress" as a result of being bullied. The human resources department is overwhelmed because of an ever-dwindling staff size due to the company's severe budget problems. Last year alone, response rates to claims fell to an all-time low of 4%: Only 4 out of every 100 serious claims led to remediation.

One of the most recent situations involved a prolonged period of bullying aimed at a shy and reserved member of the staff, Mr. Ronald Winfrey, 34. Over a period of months, one of Ronald's coworkers would make degrading comments towards him, including making fun of his weight, appearance, eating habits, and intelligence. Ronald reported the harassment to the human resources department on multiple occasions but, due to being understaffed, the department was unable to process his claims. Over time, Ronald became increasingly depressed and started feeling like he was a worthless person. Finally, after months of belittlement and degradation, Ronald attempted suicide by taking a large dose of prescription painkillers. He slipped into a coma and died two days later.

As has become typical at Hampton & Sons, Ronald's coworker was never penalized for his harassment despite strong evidence implicating him.

Generic Information

Here is the information about the company:

Thirty years ago, the Hampton & Sons Corporation was a leader in industry and manufacturing. Most Hampton employees worked on an assembly line, helping to produce automobile parts. Hampton & Sons was especially well known for their manufacture of a specific type of engine piston that helped to improve the lifetime of the

engine. Back then, only a tiny fraction of employees worked in an office, except for the managers and supervising staff.

Several years ago, however, the company has undergone a major change in manufacturing.

Since 2002, Hampton & Sons has transferred all production to a mechanical assembly line that does not require any manual labor on the part of the staff. Instead, many of the manufacturing workers take on the role of quality control, to ensure the engines are uniformly produced. Also, much of the staff now works as mechanics, ensuring that the machines are functioning properly. Currently, the majority of the employees have a bachelor's degree in mechanical engineering, allowing them to make advanced modifications to the production machines. Because of this modernization, the workers now work in office spaces and are able to make changes to the mechanical production line on their computers.

In light of these changes, Hampton & Sons remains a leader in the field for manufacturing automobile engine parts.

Curriculum Vita

Stephanie C. Cerce

Education

May 2017 (expected)	Enrolled in Psychology Ph.D. Program Concentration: Social Psychology Advisor: Michael Gill, Ph.D. Lehigh University, Bethlehem, PA
May 2011	B.A. Psychology & History, Summa Cum Laude Honors Thesis: "Evolutionary Value of Self-Deception? The Effects of Self-Deception on Ability to Deceive Others" Rutgers University, New Brunswick, NJ

Research Interests

My research interests are focused on moral decision-making, including how free will perceptions influence blame judgments.

Fellowships & Grants

2013	Awarded Dale Strohl Graduate Summer Research Fellowship

Research Experience

June 2013- April 2014	*Master's Thesis*, Advisor: Dr. Michael Gill, Lehigh University "People are Neither Compatibilists Nor Incompatibilists: They Maintain Distinct, Inconsistent Intuitions Regarding Determinism and Moral Responsibility" -Research partially funded by the Dale Strohl Graduate Summer Research Fellowship
Sept 2012- April 2014	*Research Assistant*, Advisors: Dr. Gill & Dr. Packer, Lehigh University "When and How Moral Beliefs Give Purpose to Moral Action" -Research funded by The John Templeton Foundation
Sept 2012- June 2013	*First Year Project*, Advisor: Dr. Michael Gill, Lehigh University "Breaking the Link Between Intentional Harm and Blame: The Role of Perceived Character Implantation"
June 2011- May 2012	*Research Assistant*, Advisor: Dr. David Wilder, Rutgers University
Sept. 2010- May 2011	*Senior Honors Thesis* (2 studies), Advisor: Dr. David Wilder, Rutgers University "Evolutionary Value of Self-Deception? The Effects of Self-Deception on Ability to Deceive Others"

Professional Presentations

April 2014 Brown Bag presentation on Master's Thesis
April 2013 Brown Bag presentation on First Year Research Project

Manuscripts

Gill, M. J. & Cerce, S. (in prep) He Never Willed to Will That: Historical Information about an Immoral Actor Attenuates Blame by Altering Particular Free Will Perceptions (Control of Self-Formation), While Leaving Others Intact (Freedom of Action).

Cerce, S. & Wilder, D. (under review) Evolutionary Value of Self-Deception? The Effects of Self-Deception on Ability to Deceive Others. *Psi Chi Journal of Psychological Research.*

Conference Posters

Cerce, S. & Gill, M. (March 2014) *Undermining perceptions of "second-order" free will via historical information about a transgressor disconnects perceived intentionality from blame*. Poster presented at the Eastern Psychological Association Annual Meeting, Boston, MA.

Cerce, S. & Gill, M. (February 2014) *Breaking the link between intentional harm and blame: the role of perceived character implantation*. Poster accepted to the Society for Personality and Social Psychology Annual Meeting, Austin, TX.

Cerce, S. & Wilder, D. (March 2012) *Evolutionary Value of Self-Deception? The Effects of Self-Deception on Ability to Deceive Others*. Poster presented at the Eastern Psychological Association Annual Meeting, Pittsburgh, PA.

Teaching Experience

Summer 2014 Instructor – Introduction to Psychology, Lehigh University
Prepared original course materials and lectures.

2014 Level II Certificate for Teacher Development Program, Lehigh University

2013 Graduate Teaching Assistant – Introduction to Psychology, Lehigh University
Guest lectured on original material.

2012 Graduate Teaching Assistant – Personality Psychology, Lehigh University
Guest lectured for one class period. Held review sessions for each exam. Graded essay assignment. Wrote original exam questions.

2012 Level I Certificate for Teacher Development Program, Lehigh University

2010 Peer Instructor – "Exploring Psychology", Rutgers University
Primary instructor and created original course material for a semester-long course. Facilitated class discussion. Graded all assignments.