Macalester Journal of Philosophy

Volume 19
Issue 1 *Spring 2010*

Article 7

10-7-2010

Determined to Feel Free: The Psychological Reality of Moral Responsibility

Sara Gottlieb
Macalester College

Determined to Feel Free: The Psychological Reality of Moral Responsibility

Sara Gottlieb

"My first act of free will shall be to believe in free will"
- William James

I. Introduction

Among introductory topics in philosophy, the problem of free will is arguably far more troubling to the ordinary student than many other issues. Free will and conscious control, for one, affects every single individual – there is no sense in which we can rise above concerns regarding the autonomous control of our actions. And second, the issue of free will is inextricably linked to concerns of moral responsibility. Our fear of determinism is a gaping existential anxiety that deep down we, as humans, are not in control of our own choices. So all the deliberation we perform in order to pick the correct decision is pointless, it would seem, because everything has already been preordained by the state of our brains. Steven Pinker describes the consequences of relinquishing the feeling of conscious deliberation:

> If you suffer from this anxiety, I suggest the following experiment. For the next few days, don't bother deliberating over your actions. It's a waste of time, after all; they have already been determined…No, I am not seriously suggesting you try this! But a moment's decisions should serve as a Valium for the existential anxiety. The experience of choosing is not a fiction, regardless of how

> the brain works…You cannot step outside it or let it go on without you because it *is* you. If the most ironclad form of determinism is real, you could not do anything about it anyway, because your anxiety about determinism, and how you would deal with it, would also be determined. It is the existential fear of determinism that is the real waste of time (Pinker 2008, 312).

Simply put, we blame people for an evil act or bad decision only when they intend the consequences and could have done otherwise. After all, we usually put less blame on the hunter who shot his friend thinking he was a deer than one who murders intentionally. We tend to show mercy to the victim of torture who betrays a comrade and a delirious patient who lashes out at family members because we feel they are not in command of their faculties. Again, we do not put small children on trial if they incidentally cause a death, nor do we try inanimate objects and non-human animals in the court of law – we consider them to be constitutionally incapable of making an informed choice.

Free will and moral responsibility reads as a classic paradox, based on the following set of inconsistent claims:

1. Every action is caused.
2. If an act is caused, then it is determined.
3. If an act is determined, then it is not done by free will.
4. If an act is not done by free will, then its agent is not fully morally responsible for that act.

5. If an agent is not fully morally responsible for that act, then that agent should not be punished for the act.
6. Yet some agents should be punished for particular acts.

The debate over these claims emerges as a clash primarily between compatibilists and incompatibilists. This disagreement relies mostly upon claims 3-4. Incompatibilists assert that if an act is completely determined, then the agent is not fully morally responsible for that act. Compatibilists, on the other hand, deny this assertion in favor of preserving some degree of responsibility even when acts are determined.

The worry of free will comes under scrutiny in light of the scientific image. Owen Flanagan, in his book *The Problem of the Soul*, attempts to reconcile the manifest image with the scientific image of the mind. His portrayal of the problem of free will as elucidated by the scientific image will be the focus from here on out:

> Suppose I accept that everything that happens has a set of causes sufficient to produce it. Given this set of causes, whatever they are, this effect, whatever it is, is necessary. Suppose I also accept that when I act I do so for reasons, often after I think and deliberate. Such acts are free. But when I act freely, we can ask, why did I think, deliberate, and reason in the way I did rather than in some other way? […] There is no denying that I reason, think, and deliberate; what is denied is that my reasoning, thinking, and deliberating are self-originating, things that

> cause my actions without themselves having causes that make them necessary. If, however, I think of my action in terms of the wider causal nexus, as falling under the law of universal causal determinism, then I can't conceive of my action as having been different than it was (Flanagan 2002, 45).

As Pinker explains, people hoping that an uncaused soul might rescue a sense of personal responsibility from the deterministic natural sciences, mind sciences, and evolutionary psychology, are in for disappointment. Daniel Dennett, after all, points out that the last thing we want in a soul is freedom to do *anything* that it desires (Dennett 1984). If behavior were driven by an utterly *free* (in all sense of the term) will, then we really could not hold people responsible for their actions since an individual would not be deterred by the threat of punishment or feelings of guilt. In short, a completely free agent, floating on a different plane from the powers of cause and effect, would be unaffected by moral codes of law and ethics and not have any reason to act since its will would be completely unrestrained (Pinker 2008).

This paper argues for a naturalist account of the problem of free will. I will focus on the *feeling* of being a morally responsible agent, not the *ontological* truth-value or whether or not human beings *are* actual morally responsible agents. Again, this focus on the *feeling* is a second-order question that does not claim to answer the first-order problem of whether any of the terms to which I will refer – compatibilism, incompatibilism, free will, determinism, etcetera - are true in their strict logical sense. I aim to show that the *feeling* of being a morally free and responsible agent

is a facet of human cognition that is not readily undermined by contradictory knowledge of a naturalistic universe. My naturalized account is similar to that of Hume's skeptical approach to the problem of induction, and appeals to evidence from both psychology and experimental philosophy. I call special attention to affective emotional states, providing evidence for the profound influence of emotion on moral decision-making, and the divide between abstract (i.e., philosophically guided) and concrete (i.e., emotionally laden) intuitions.

II. Does knowledge of a naturalistically determined universe lead to immoral behavior?

Pinker espoused an anecdotal account of why we should fear determinism, but social psychologists have approached the issue from an empirical perspective. In a recent investigation, Kathleen Vohs and Jonathan Schooler asked whether moral behavior draws on a belief in free will. They examined whether inducing participants to believe that human behavior is predetermined would encourage cheating. Participants read one of two excerpts from Francis Crick's *The Astonishing Hypothesis*; those in the experimental condition read about a fiercely deterministic and reductionist view of the brain: "You, your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules. Who you are is nothing but a pack of neurons." The control group read an excerpt that discussed consciousness but did not mention free will. Participants then engaged in a seemingly unrelated task, and were told to take a particular amount of money as compensation for their participation after the experimenter had left the room.

However, more money was available so that the participants could cheat the system by taking more than instructed. They found that participants who read deterministic statements cheated by overpaying themselves for performance on a cognitive task – a type of immoral behavior. Thus, they suggest that the debate over free will has societal, as well as scientific and theoretical implications. The authors write:

> If exposure to deterministic messages increases the likelihood of unethical actions, then identifying approaches for insulating the public against this danger becomes imperative. Ultimately, in order to oppose the unfavorable consequences of deterministic sentiments, the field must first develop a deeper understanding of why dismissing free will leads to amoral behaviors. Does the belief that forces outside the self determine behavior drain the motivation to resist the temptation to cheat, inducing a 'why bother' mentality? Doubting one's free will may undermine the sense of self as agent (Vohs and Schooler 2008, 54).

In a follow-up study by noted social psychologist Roy Baumeister, he and his colleagues sought to extend the results of Vohs and Schooler into a broader context, namely helping and aggression. They hypothesized that a disbelief in free will increases aggression and reduces helping since free will is crucial for motivating people to control their automatic impulses in favor of more prosocial types of behavior. Vohs and Schooler proposed that doubting free will serves as a subtle cue that exerting volition is futile and induces the "why bother" mentality. Thus, Baumeister

sought to demonstrate that making people disbelieve in free will acts as a nonconscious prime to act in relatively automatic ways, which includes acting upon more selfish impulses rather than exerting control and self-restraint.

In their 2009 study, Baumeister, Masicampo and DeWall used an improved methodology (most notably, improved prime passages describing free will or determinism) built upon Vohs and Schooler's experiment and found that prosocial tendencies (as characterized by a willingness to help) were reduced among participants who were induced to believe in determinism and disbelieve in free will. That is, these participants were less willing to help across a wide array of situations and opportunities than were those in the control group (i.e., those who were induced to believe in free will). In follow-up experiments, they focused on self-control and found that a disbelief in free will makes people reluctant to expend their energy in acts of self-restraint. Furthermore, since aggression is an automatic and antisocial response, they hypothesized that the deterministic beliefs are related to aggression by depleting the ability to override certain types of socially undesirable automatic impulses. Specifically, they found that those individuals who were chronically high in disbelief in free will were also less likely to help others in distress than those who were skeptical of rejecting free will.

The broader implications of these results suggest that volition and self-control require the person to expend energy, and that these expenditures enable them to act prosocially. Apparently, a disbelief in free will subtly reduces people's ability to expend that energy, and the disbelief in free will serves as a cue to act on impulse, which generally promotes selfish actions such as aggressing and refusing to help. Baumeister et al. concede that these results

have nothing to say about free will as objective reality; indeed, their results could hold valid even if free will is a complete illusion. The authors write, "Some philosophical analyses may conclude that a fatalistic determinism is compatible with highly ethical behavior, but the present results suggest that many laypersons do not yet appreciate that possibility" (Baumeister et al. 2009, 267).

Without explicit statement, both Vohs and Schooler and Baumeister et al. seem to be showing that people hold incompatibilist views regarding determinism and moral responsibility. More generally, they hold that a belief in free will, regardless of its ontological truth, is pertinent to prosocial behavior and morality. These experimenters sought to answer the question of whether the knowledge of determinism undermines a sense of moral agency, and they adhere to the conclusion that such reminders negatively affect our social interactions in highly unconscious, automatic ways. But there also exists a competing view that this feeling of being a morally free and responsible agent, illusory or not, has limits to the capacity by which it can be undermined. Daniel Wegner, a Harvard psychologist and author of *The Illusion of Conscious Will*, writes:

> I'm a case in point. I've devoted years of my life to the study of conscious will…If the illusion could be dispelled by explanation, I should be some kind of robot by now, a victim of my own nefarious schemes. No self, no magic, no inner agent. Yes, it's true, when I'm on the dance floor I may look a bit robotic to some – but I'm happy to report that despite my personal flurry of illusion busting, I remain every bit as susceptible to the experience of conscious will as the next

> person. It feels like I'm doing things (Wenger 2008, 237).

I must clarify that I do not doubt the view that knowledge can diminish our sense of moral responsibility and conscious control by varying degrees. A greater understanding of the biological bases of mental illness has indeed caused subtle changes to the ways in which we ascribe punishment to criminals whose actions were deemed *beyond their control*. But in the minute-to-minute everyday sense – somebody cuts me off while changing lanes on the highway, or misses a scheduled meeting and leaves me waiting – I argue that the ways in which we see both ourselves and others as moral agents *cannot* be undermined. A Humean approach appropriately illustrates the naturalist account for which I am arguing.

III. Hume's skepticism: the problem of induction and the problem of free will

Hume distinctively sought to ground human knowledge on empirical natural facts to develop a naturalized account of human cognition. Hume describes the problem of induction by the fact that experience does not produce the idea of an effect from an impression of its cause by reason, but rather by the imagination or certain association and relation of perceptions. The crux of the problem is that if understanding produced inductive conclusions, then inductive reason would be based on the assumption that nature is uniform, or that those instances that we have not experienced follow the experiences of the past. If this were the premise to be established by reasoning, then reasoning would be either deductive or probabilistic. Since the principle cannot be proven deductively, since

whatever is proven deductively is a necessary truth, and this principle isn't necessary since its antecedent is consistent with the denial of its consequent. Causal or probabilistic reasoning also cannot prove it, because it is presupposed by all such reasoning.

In *The Enquiry Concerning Human Understanding*, Hume classifies all reason into either *relation of ideas* or *matters of facts*. Matters of fact are propositions whose negations do not necessarily lead to contradictions. Hume regards my assertion that the sun *will not* rise tomorrow no less intelligible, nor no more a contraction, than the assertion that it *will*. So all reasoning concerning matter of fact is founded on the relation between cause and effect; after repeated experiences of events of type A being followed by type B, we will be prone to infer event B following an experience of event A. There is no necessary connection between events A and B, and the post-hoc reasoning that B is followed by a fact A is a type of inductive reasoning. This causal chain of reasoning extends backward into the further past as we infer *a posteriori*. But we cannot establish a connection without having had an experience in which event A is followed by event B, so it logically follows that cause and effect cannot be reasoned for using *a priori* argumentation (Hume 1999).

This is the negative component of Hume's argument, as he rules out accounts of induction that view it as the product of reason. Causal reasoning is the strongest associative relation and perhaps also the most important, since by means of inferring this relation can we go beyond the evidence of our memory and senses. We may even say that we are, in a sense, "addicted" to inferring a connection between the present fact and that which is inferred from it. Causal reasoning is not *a priori*. We discover causes and

effects through experience, not by reason, and most ordinary causal judgments become so familiar over time that our judgments seems automatic.

The process that produces our causal expectations is itself causal; customs and habits determine the mind to assume that the future is congruent with the past insofar as it assumes like effects to follow from like causes. Hume does not go so far as saying that we are justified when we engage in inductive reasoning, yet also realizes that his negative phase exhausts all other forms of reasoning. We have no choice but to induce, and we have no choice but to expect laws of the past to hold similarly in the future. I have no logical reason to believe that the sun will rise tomorrow, but I just cannot help it. When I get out of bed in the morning, I have no reason to believe that the gravity of today is the same of the gravity of yesterday, and that I won't fly toward the ceiling. When I open my mouth to speak, there is no guarantee that words will come forth. Yet without any of these inductive inferences, I could not exist. I could not convey ideas through language, and I could not move my body. Without inductive properties, it seems that human existence comes to a complete halt.

This skeptical approach parallels the problem of free will. Vohs and Schooler concluded that we must insulate the public from knowledge of determinism since it leads to an increase in morally wrong behaviors. But if Hume is right with regard to inductive inference, then we may view the undeniable feeling of being a free and morally responsible agent as an analogously similar and necessity facet of naturalized human cognition. If we imagine a world in which we believe ourselves to be completely constrained, and within a world in which we have no ability to do otherwise and all actions are determined by antecedent

causes, which are themselves determined by antecedent causes, so on and so forth, then even thinking about our own freedom, as we are presently doing, becomes exceedingly problematic. Our present thoughts are a determined act that none of us are deliberatively and consciously willing; we must not doubt that we are indeed reasoning at the present moment, but this reasoning is nonetheless derived from antecedent causes beyond our control. Conceiving this world feels beyond the capacities of human cognition, and navigating a world in which we come to truly internalize this sense of naturalistic determinism, although logically possible, appears far too abstract for the human mind.

By taking the Humean route, I argue that, in light of a strong scientific image, we indeed have reason to believe that actions have antecedent physical causes. These actions may not have one sufficient or necessary cause (i.e., monocausation must not necessarily hold true), but actions are nonetheless driven by complex sets of factors that necessarily guide the future in one direction. Hume argued that we have no justified reason to believe that the sun either will rise tomorrow; in parallel fashion, we have no reason to believe both ourselves and others are the originators of actions and should *feel* morally responsible as such. Inductive inference is central to the very reasoning that makes us human, and I argue that the feeling of being morally responsible, and the feeling that other people are morally responsible as well, is naturalistically grounded in a human necessity that can not be undermined by contradictory evidence.[1]

---

[1] As I stressed earlier, I do not doubt the degree to which ascribing responsibility to both ourselves and others can be diminished in varying degrees, unlike the *absolute* necessity of inductive inference. However, there exists a sincere limit to which a feeling of agency can be

IV. Are we intuitive compatibilists or incompatibilists?

The project of experimental philosophy is, in part, to explore the common intuitions of normal folk. This innovative movement seeks to return the discipline of philosophy to a focus on questions about how people actually think and feel; it departs from a long-standing tradition of armchair exploration as experimental philosophers go out and conduct systematic *experiments* to reach a better understanding of people's ordinary intuitions about philosophically significant questions. Experimental philosophers approach the question of free will by questioning how the paradox of responsibility arises and why it persists both among professional philosophers and the folk. If only accomplished philosophers had conflicting intuitions about the topic, then the paradox would not be so troubling or persistent in the progression of literature. And when experimental philosophers speak of intuition, they refer to inclinations to believe claims whose attractiveness is not dependent on any conscious inference. Common intuitions are not especially reflective, and these are the ones studied most by experimental philosophers.

Shaun Nichols, a leader in the field, explored the lay understanding of choice, and found that both children and adults tended to treat moral choices as indeterminist. Here, they were given cases of moral choice events (e.g., a girl steals a candy bar) and physical events (e.g., a pot of water comes to a boil) and asked whether, if everything in the world was the same right up until he event occurred, the event must occur. Both the children and the adults were more likely to answer yes in the case of the physical event

---

undermined, and for this reason the analogy between induction and free will is highly appropriate, although not completely parallel.

than the moral event (Nichols 2004). Thus, these results seem to suggest that people believe that some cases of moral decisions are not determined. People *intuitively* believe that their actions are not governed by the same physical laws that govern inanimate objects.

A parallel agenda recently explored whether people are intuitive compatibilists or incompatibilists about moral responsibility. Woolfolk, Doris and Darley provided participants with a story about an individual who is kidnapped and given a "compliance drug" that makes it impossible for him to disobey orders. When the kidnappers order him to act immorally, he has no choice but to obey. Participants were either in the low identification condition and told that the agent did not want to perform the immoral act, or in the identification condition, and told that that he wanted to perform it all along. The results demonstrated a main effect of identification; those in the high identification condition attributed significantly higher ratings of responsibility to the agent that those in the low identification condition (Woolfolk et al. 2006). This is in line with Frankfurt's compatibilist claim that responsibility depends on identification (Frankfurt 1969).

A further study by Nahmias, Morris, Nadelhoffer and Turner poses a more significant threat for the view that people are intuitive incompatibilists. They found that participants hold agents morally responsible even when that agent is acting in a completely deterministic universe. Participants read the following scenario

> Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these 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. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 pm on January 26th, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 pm on January 26th, 2195.

After reading the vignette, participants were immediately asked whether or not Jeremy is morally responsible for robbing the bank. Strikingly, 83% responded that he is blameworthy. This effect was replicated in two further experiments with different scenarios, providing strong evidence that people regard moral responsibility compatible with determinism (Nahmias et al 2005).

The results described above raise a significant dilemma: if people so consistently demonstrate compatibilist positions on experimental questionnaires, then for what reason does the debate over incompatibilism persist? Nichols and Knobe suspect that there is more at work here than has typically been acknowledged. In their view, most people have a logical bias toward incompatibilist theories of moral responsibility (at least in our culture), and these tendencies are easily elicited in philosophical discussion and especially within the classroom. However, in addition to these preferences, people also exhibit immediate affective

reactions to vignettes describing immoral behaviors. They write, “What we see in the results of Nahmias and colleagues is, in part, the effect of these affective reactions. To uncover people’s underlying theories, we need to offer them questions that call for more abstract, theoretical cognition” (Nichols and Knobe 2008, 109).

Nichols and Knobe wished to explore whether participants would be more likely to report compatibilist or incompatibilist intuitions if they controlled for emotional and motivational factors. Thus, they isolated two conditions: in the *concrete* condition, they aimed to elicit greater affective responses, whereas in the *abstract* condition, they aimed to trigger theoretical, armchair cognition. They hypothesized that people would be more likely to act as incompatibilists in the concrete (i.e., emotional) condition.

Participants were provided with descriptions of Universe A and Universe B; in Universe A, everything that happens is completely caused by whatever happened before it, and this is true from the very beginning of the universe up until the present. Universe B describes an environment in which *almost* everything is completely caused by whatever happened previously. Thus, the key difference between the two is that, given the past, whether or not each decision has to happen. Participants were asked which universe is most like their own, and to explain their answer. Over 90% reported to live in a universe most like Universe B, suggesting that people intuitively see their world as indeterministic.

As a follow-up, participants then answered a question regarding compatibilism. They were randomly assigned to either the *concrete* condition or the *abstract* condition. In the concrete condition, they read the following statement:

> In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and three children. He knows that it is impossible to escape from his house in the even of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family. Is Bill fully morally responsible for killing his wife and children?

In this condition, 72% gave a compatibilist response, indicating that Bill should be held fully morally responsible. This is synonymous with the findings of Nahmias et al. 2005). In the abstract condition, they were asked the following: In Universe A, is it possible for a person to be fully morally responsible for their actions? As hypothesized, an incredible 86% of participants gave the *incompatibilist* response. In short, these results indicate that affect interacts with judgments to elicit compatibilist intuition.

This experiment still leaves open questions about whether the responses are actually the product of emotion. It is conceivable that what mattered was the concreteness itself, not the actual affect associated with concreteness. That is, perhaps the compatibilist responses were elicited because the scenario described a particular act by a particular individual, and the results are not due to affect alone. To investigate this, Nichols and Knobe created a new moderator in which subjects were randomly assigned to either the *high affect* or the *low affect* condition. For high affect, they were asked: As he has done many times in the past, Bill stalks and rapes a stranger. Is it possible that Bill is fully morally responsible for raping the stranger? In the low

affect condition, they read: As he has done many times in the past, Mark arranges to cheat on his taxes. Is it possible that Mark is fully morally responsible for cheating on his taxes? For half of the subjects, the question stipulated that the agent was in Universe A; for the other half, it was in Universe B. Thus, they created a 2x2 factorial design of affect and universe. They found that even when concreteness was controlled for, affect impacts people's intuitions about responsibility under deterministic circumstances (Nichols and Knobe 2008). The complete results were as follows:

	Agent in indeterminist universe	Agent in determinist universe
High affect case	95%	64%
Low affect case	89%	23%

Overall, we find that both affect and concreteness are important in the process that generates compatibilist judgments. Explanatory models exist for both of these criteria as independent identities, but as we will see in the coming section, they complement each other remarkably well as I come to argue for a naturalized account of treating both ourselves and others as morally free and responsible agents.

V. Explanatory models: Affect, social intuitionism, and concreteness

Within psychology, the field of social cognition has recently shifted focus from the deliberative aspects of decision making to the highly automatic and unconscious ones. More specifically, emotion research portrays automatic processes as heavily dependent upon affective states. The primary method with which psychologists test for nonconscious influences on cognition is called *nonconscious priming*. John Bargh and colleagues asked participants to complete a scrambled sentence task either loaded with terms associated with the elderly or neutral terms. Unbeknownst to the participants, researchers measured the speed with which they moved about the room, and found that those who were nonconsciously primed to think about the elderly actually walked slower than those in the control condition. They showed that even when a goal or concept is induced implicitly, an effect may still emerge. In a follow-up study, Bargh et al. provided participants with scrambled words either loaded with the concept of rudeness of politeness. They were instructed to notify the experimenter when they were ready to move to the next task; however, the experimenter was engaged in conversation and the participant would have to interrupt in order to give notification. The experimenters found that those in the rudeness condition interrupted significantly more than those in the other condition, remarkably demonstrating that our decisions are influenced in ways that fall far outside of awareness (Bargh et al. 1996). They write, “For every psychological effect (e.g., behavior, emotion, judgment, memory, perception), there exists a set of causes or antecedent conditions that uniquely lead to that effect” (Bargh and Ferguson 2000, 295).

Social psychologist Jennifer Lerner has heavily influenced the literature on emotional cognition, and found that when subjects' were primed to arouse negative emotions, they held agents more responsible and more deserving of punishment; indeed, this even held true when the negative emotions were aroused by an unrelated event (Lerner et al. 1998). In their experimental design, participants in an anger condition watched a violent film of a bully beating up a teenager, while subjects in the neutral-control condition watched a video clip of abstract figures. All participants then attributed responsibility for various negligent behaviors to fictional characters. Interestingly, those in the anger condition gave higher responsibility ratings than subjects in the control condition. This provides significant evidence that emotions induced by outside events (i.e., the film) impact responsibility judgments in clearly unrelated scenarios. A natural interpretation of this data is that the emotion biased the participants in making their assessments of reasonability, and thus detracted from their abilities to reason in a more logical manner.

This literature built a remarkably strong foundation for the emerging field of moral psychology, which investigates the role of emotion, automatic judgments and other motivational factors in forming intuitions in ethically relevant situations. Social psychologist Jonathan Haidt has pioneered morality research with his *social intuitionist model* of moral judgments (Haidt 2001). His central claim states that moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, ex post facto moral reasoning; this is not an anti-rationalist model, but rather alludes to the complex and dynamic ways by which intuition, social influences and reasoning interact to produce moral judgment.

Haidt motivated his model with what he calls *moral dumbfounding.* When participants were presented with a story describing a brother and sister (Julie and Mark) who decide to make love but have safe sex and actually grow closer from the experience, most responded that they felt it was wrong but could not articulate why. Haidt explains that while moral emotions (e.g., sympathy, disgust) are sometimes significant inputs in the reasoning process, they are not the direct causes of moral judgments. Participants who read about Mark and Julie experienced a strong visceral response to the vignette, and felt a small flash of disapproval for the act before they were able to actually articulate why they felt this way. In line with the social intuitionist model, these individuals made quick, automatic moral intuitions as influenced by their affective states, and then attempted to follow these intuitions with deliberative reasoning as justification for their responses. At a certain point, they realized that their intuitions were logically irrational; when asked, after many times, why the act was wrong, they ultimately responded with, "I don't know why, but it *just is*."

Haidt's social intuitionist model provides plausible explanation for the different intuitions generated in Nichols and Knobe's high and low affect cases. But there remains the case of judgments made in concrete versus abstract situations, and Haidt's research in emotion and moral intuition appropriately leads to this discussion. Concrete situations aptly describe cases occurring in our own worlds, with which we are highly familiar, and from which we can draw a great deal of experiential knowledge. Abstract conditions, however, concern a world with which we are far less familiar, and must draw upon logical a priori reasoning to form judgments.

Concreteness and abstractness are clearly contrasted in the case of Nichols and Knobe's Universe A and Universe B. Recall that the vast majority of participants deemed Universe B most like our own; thus, when answering questions about Universe B, they drew upon concrete notions of everyday experience to form intuitions. Universe A, however, is slightly ambiguous since participants did not believe themselves to have first-hand knowledge of and experience with a world in which human decisions are completely governed and determined by the same physical laws that govern inanimate objects. Do the participants have family in this alternate universe? What are the environmental influences on action? Is there law enforcement? In this respect, Universe A is far less concrete. The majority of people believed themselves to be familiar with Universe B, and hence formed a vast array of beliefs, images and other cognitive attitudes about this world. These complex sets of formations draw upon affective responses, and form distinctive cognitive attitudes about the world. We may conclude from this discussion that concreteness tends to stimulate compatibilist intuitions and abstractness tends to stimulate incompatibilist intuitions.

There is markedly distinctive manner by which people form intuitions as they move from the philosopher's armchair or an academic classroom and into the concrete nature of real world social interactions. Haidt's social intuitionist model of moral judgment motivates this assertion, as concrete situations draw upon dynamic sets of experience and knowledge; since an astonishing proportion of our memories and values are stored as involving strong emotional components, concrete situations elicit the automatic and unconscious intuitions for which the model argues. Abstract cases, on the other hand, elicit a separate

type of intuition. The professional philosopher, or simply the law person who is asked to reason for an abstract case about some alternate universe, allows for more of the post hoc reasoning that Haidt includes in his model. Since moral intuitions are, under many circumstances, tiny flashes of either approval or disapproval, there are strong evolutionary reasons for the persistence of intuitive judgments in highly concrete cases. The disgust emotion, for example, originated as an oral defensive mechanism against noxious stimuli and parasitic substances, yet has transitioned far from its evolutionarily intended purpose and into the moral domain. Inbar, Pizarro, Knobe and Bloom found that an individual sensitivity to physical elicitors of disgust predicts an intuitive disapproval of homosexual behavior, which is potentially regarded as a type of sexual impurity (Inbar et al. 2009). The disgust response has a strong evolutionary value, yet causes an over-generalized response in some present moment social situations. But people are nonetheless able to override implicit judgments of this sort and rely upon post hoc deliberation to form an explicit preference. Applied to the discussion of concrete and abstract situations, we know that people are less able to regulate their cognitive processes, and instead fall back on more automatic judgments, when under strong emotional influence. Concrete situations elicit strong, often visceral, emotional states that cause people to rely on certain automatic intuitions that do not necessarily coincide with the intuitions they would form if they had ample time and cognitive resources to devote for full deliberative reasoning.

In this section, I examined the cognitive effects of affect and concreteness on intuition. High affect and concrete situations elicit compatibilist intuitions more than low affect abstract situations. This is demonstrated by

experimental philosophers, and is motivated by Haidt's social intuitionist model.

V. Conclusion

This paper addresses the problem of free will from a psychological perspective. I asked whether knowledge of determinism undermines the *feeling* of being a morally responsible agent. Vohs and Schooler used experimental investigation to find that people primed to believe in a naturalistically determined universe acted more immorally than those in a control condition. If their conclusion is correct, then it seems as if we have reason to worry. However, I argue here for a naturalized account of referring to both ourselves and others as morally free and responsible agents. I draw the analogy to Hume's skeptical account of the problem of induction; this feeling of moral agency is fundamental to and extricable from human cognition, much in the same way that inductive inference presupposes human thought and interaction. The problem of free will is perhaps so troubling because conflicting intuitions regarding the compatibility or incompatibility of moral responsibility are widespread and persistent among both the philosopher and lay folk. Experimental philosophers have devoted empirical investigation to the issue at hand by uncovering intuitions of compatibilism or incompatibilism and varying affective elicitors and the concreteness of target situations. These factors accurately account for conflicting intuitions, as demonstrated by Haidt's social intuitionist model of moral judgment. When our intuitions refer to high affect and concrete cases, we fall back on automatic intuitions in favor of compatibilist judgments. As we leave the armchair and venture into the complex social dynamic of real-world situations, these intuitions are fundamental to human

cognition. This is a naturalized account of moral agency, for we see that knowledge of naturalism or determinism does not readily undermine the way we treat both ourselves and others as morally free and responsible agents. Even in a world believed to be naturally determined, we see moral action as unrestrained by the physical laws that govern inanimate objects. Ontological dispute surrounds the first-order truth or free will or determinism, but the psychological truth of *feeling* morally free is an undeniable reality.

Bibliography

Bargh JA, M Chen, and L Burrows. 1996. Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology.* 71:230-44.

Bargh JA and MJ Ferguson. 2000. Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin.* 126:925-45.

Baumeister, R, EJ Masicampo, and C DeWall. 2009. Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin.* 35:260-268.

Dennett, DC. 1984. *Elbow room: the varieties of free will worth wanting*. Cambridge, Mass: MIT Press.

Flanagan, OJ 2002. *The problem of the soul: two visions of mind and how to reconcile them*. New York: Basic Books.

Frankfurt, HG 1969. Alternate possibilities and moral responsibility. *The Journal of Philosophy.* 66:829-839.

Haidt J. 2001. "The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review.* 108:814-34.

Hume, D and TL Beauchamp. 1999. *An enquiry concerning human understanding*. Oxford philosophical texts. Oxford: Oxford University Press.

Lerner, JS, JH Goldberg, and PE Tetlock. 1998. Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility". *Personality & Social Psychology Bulletin.* 24:563.

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

Nichols, S. 2004. The folk psychology of free will: Fits and starts. *Mind and Language*. 19:473-502.

Nichols, S and J Knobe. 2008. Moral responsibility and determinism: The cognitive science of folk

intuitions. In *Experimental Philosophy*, eds. Knobe and Nichols, 105-126. New York, NY: Oxford.

Pinker, S. 2008. The fear of determinism. In *Are we free? Psychology and free will*, eds. Baer, Kaufman and Baumeister, 311-324. New York, NY: Oxford.

Vohs, KD and JW Schooler. 2008. The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science* 19:49-54.

Wegner, DM 2002. *The illusion of conscious will*. Cambridge, Mass: MIT Press.

Woolfolk, R., J. Doris and J. Darley. 2006. Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*. 100:283-306.