Stubborn Moralism and Freedom of the Will

David A. Pizarro and Erik G. Helzer

Cornell University

To be included in:

Baumeister, R.F., A.R. Mele, and K. D. Vohs (Eds.) Free will and consciousness: How might they work? Oxford University Press.

Stubborn Moralism and Freedom of the Will

Imagine discovering that your neighbor, whom you have known for years, is in fact a very sophisticated robot that has had all of his behavior programmed in advance by a team of robotics experts. This information might cause you to re-evaluate all of your interactions with him. Where you previously may have become angry with him over the slightest offense, you may now feel the need to remind yourself that he is merely following his programming; he has no say over his own behavior. Likewise, you might find it unreasonable to hold him in contempt for having beliefs that conflict with your own. In short, you may find yourself wanting to suspend your moral evaluations about his beliefs, judgments, and behaviors across the board because, after all, he did not act freely.

This connection between freedom and moral responsibility is evident in the long and storied debate over free will among philosophers, many of whom have argued that if free will does not exist, then the ability to hold individuals responsible is compromised (e.g., Clarke, 2003). On this view, in order to hold an individual morally responsible for an act, the possibility must exist that he could have done otherwise. The deterministic processes that give rise to the behavior of the robot neighbor would, of course, be incompatible with this requirement—it is simply a fact that the robot lacks the ability to do anything other than what he was programmed to do. Some have questioned this conclusion, arguing that even if human beings are not free in this ultimate sense (and that humans are all simply very complex, organic versions of the robot-neighbor), the ability to hold others morally responsible remains logically unaffected (Pereboom, 2001; Strawson, 1974). Nonetheless, the deep connection between freedom and moral

responsibility remains, as disagreements are generally not over whether freedom is necessary for moral responsibility but rather over the specific kind of freedom necessary (Dennett, 1984).

In an important sense, these questions about free will and moral responsibility are beyond the scope of empirical investigation—an experiment cannot settle the question of whether moral responsibility actually requires libertarian freedom or whether determinism and moral responsibility can comfortably coexist. But individuals entirely unconcerned with scholarly debates about freedom do make judgments about moral responsibility fairly frequently, and often these judgments have serious consequences (such as social exclusion, imprisonment, and even death). The descriptive question of how people go about making these judgments of moral responsibility has therefore been of much interest to psychologists. Accordingly, a great deal of research has been conducted documenting the rules individuals seem to use when determining whether or not to hold others morally responsible. Perhaps unsurprisingly, this research has shown that people seem to care about whether an act was committed under conditions that seem to limit an agent's freedom (such as involuntary acts or accidents). Most influential theories of blame and responsibility within psychology have therefore characterized freedom as an important prerequisite for the attribution of moral responsibility, much like the normative theories of blame and responsibility from philosophy and law that influenced them (e.g., Shaver, 1985; Weiner, 1995). On these accounts, when individuals are faced with a moral infraction, they first set out to determine if a person acted freely then proceed to use that information as input into their judgment of whether to hold the person responsible.

We will argue that this view is mistaken. Rather than serve as a prerequisite for moral responsibility, judgments of freedom often seem to serve the purpose of justifying judgments of responsibility and blame. This seems true not just for judgments of "ultimate" metaphysical

freedom (i.e., freedom from determinism), but also true for freedom in the more local "agentic" sense that is central to psychological and legal theories of responsibility (such as whether or not an act was intentional or controllable).

One reason for this is that people are fundamentally motivated to evaluate the moral actions of others, to hold them responsible for these acts, and to punish them for moral violations—they are "stubborn moralists." (Although morality can refer to a very large set of behaviors and judgments, when we refer to "morality" throughout the paper, we are limiting our the definition to these aforementioned aspects of human morality—that of judging acts as morally right or wrong, and judging others as responsible for moral violations or virtuous acts.) This motivation most likely has its roots in the evolutionary forces that shaped human morality and has psychological primacy over more fine-grained social judgments, such as judgments about whether or not an act was committed freely. As such, the motivation to seek out and punish wrongdoers can push individuals in the direction of holding people responsible even when the wrongdoers do not meet the criteria for freedom required by normative accounts of responsibility. Simply put, when people say that someone acted freely, they may be saying little more than that the person is blameworthy. In support of this general argument, we first defend the claim that people are stubborn moralists before reviewing recent empirical evidence that this moralism drives judgments of freedom rather than the other way around.

Stubborn Moralism

Moral judgment comes naturally to human beings. People write stories about good and evil, make long lists of morally forbidden acts, and take a keen interest in the moral missteps of complete strangers. People are also very "promiscuous" with their moral judgments—they

readily offer moral evaluations of fictional characters, animals, and even computers. In short, while it may be true that specific moral beliefs vary across time and place, the basic belief that some acts are morally forbidden, and that people should be held responsible for these acts, appears to be a common human trait.

As evidence of the strength of these moral beliefs, consider briefly some of the putative threats that might be expected to shake confidence in morality—atheism, relativism (of the sort arising from moral diversity), and determinism. The wide dissemination of these ideas does not seem to have dented the belief that some acts are wrong, and that individuals are morally responsible for their actions (Roskies, 2006; (<XR>Chapter 10</XR>, this volume). For instance, contrary to the belief that morality hinges on the existence of God (as Dostoevsky's Ivan Karamazov claimed, "if there is no God, everything is permitted"), atheists seem to have no trouble holding people morally accountable. Nor has the knowledge that there is wide diversity in many moral beliefs seemed to undermine laypeople's belief in their own ethical systems or their belief that individuals should be held responsible for moral infractions more generally. If anything, people respond in a particularly harsh manner when presented with others who hold moral beliefs that diverge from their own (Haidt, Rosenberg, & Hom, 2003). Finally, what of the threat of determinism? If it turns out to be true that humans are all preprogrammed automatons, doesn't this invalidate the ability to hold others responsible (Greene & Cohen, 2004)? While we will address this concern specifically a bit later, Roskies (2006) has argued convincingly that neither the old threat of determinism "from above" (i.e., theistic determinism—that God's foreknowledge undermines human freedom) nor the newer threat "from below" (i.e., scientific determinism) appears to have had a wide influence on the belief that some things are wrong, that most people are able to choose between right and wrong, and that individuals deserve to be

blamed for these wrong actions (although the threat of determinism does appear to influence our own moral behavior, hinting at the possibility that different processes govern our moral self-evaluations than our evaluation of others; Baumeister, Masicampo, & Dewall, 2009; Schooler, Chapter 12, this volume; Vohs & Schooler, 2008).

The fact that people cling tightly to moral beliefs despite cultural and historical shifts that might encourage their abandonment speaks against the once-common view that morality is a thin layer masking humans' deeply selfish and amoral nature (DeWaal, 2006). On the contrary, the fact that morality thrives despite these forces suggests that the mechanisms that give rise to people's basic moral intuitions (such as the belief that people should be held responsible for their acts) are too deeply entrenched in the human mind to be abandoned easily (Haidt & Joseph, 2004). This "deep" view of morality has been increasingly bolstered by research in game theory, evolutionary biology, and economics showing that the presence of basic moral behaviors (e.g., altruistic acts) and judgments (e.g., a preference for fairness) is not inconsistent with the "selfish gene" approach to evolution by natural selection, and that evolution may actually favor individuals who exhibit such behaviors and judgments (Axelrod & Hamilton, 1981; Trivers, 1971). For instance, the most plausible account of how human altruism emerged relies on the dual mechanisms of kin selection (a willingness to act altruistically toward members of one's immediate gene pool) and reciprocal altruism (a willingness to act for the benefit of others when there is a chance that the organism will be paid in kind). Together, these mechanisms most likely gave rise to the sorts of moral emotions that proximally motivate moral action, such as empathy for the suffering of others and anger over being cheated (Frank, 1988). More recently it has even been proposed that sexual selection may have played a significant role in the preservation of morality by favoring the presence of many traits considered to be morally virtuous (Miller,

2007). For instance, men who acted virtuously (e.g., with bravery and trustworthiness) were more likely to be sought after by women either because such acts provided a direct cue to the men's fitness (indicating a higher likelihood that they would stay to help rear the offspring, thus ensuring the spreading of their genes to the next generation), or because the virtues were reliable correlates with other fitness-related cues. This bridging of morality and evolution represents one of the most significant advances in the psychology of morality, as the belief that evolution could only favor selfish organisms implies a view of morality as superficial in the sense that it is entirely dependent on cultural transmission and a proper upbringing (and by extension, that a change in culture or upbringing could eliminate human morality in one generation)—a picture of morality that is increasingly viewed as untenable.

This view of morality as deeply rooted in basic human nature finds additional support from psychologists working across various fields, including social psychology, developmental psychology, and social/cognitive neuroscience. Consistent with the view that humans are "hardwired" to be ethical creatures, recent neuroimaging studies have demonstrated that the brain systems associated with reward and distress are involved in a basic facet of human morality—a preference for fairness (a preference that not only emerges early in development, but that also seems present in humanity's close evolutionary relatives; Brosnan & DeWaal, 2003). In these imaging studies, researchers relied on a common method used to study fairness in a laboratory setting in which pairs of individuals participate in an economic game known as the "ultimatum game." In this game, one player (the "donor") is given a sum of money and told that she may allocate as much of the money as she would like to a "recipient." The recipient, in turn, must then decide whether to accept the offer (in which case, both players keep the money) or to reject the offer (in which case, neither player keeps any money). In order to be able to keep the

money, the donor must strategically allocate a quantity that will entice the recipient to accept the offer. As it turns out, a majority of individuals in the role of recipient will reject an offer that is substantially below the "fair" mark of 50% (despite the fact that accepting an unfair offer always guarantees more money than the rejection payoff of zero). Many economists have used this simple finding that individuals will take a monetary "hit" in order to punish unfair behavior as an argument for the primacy of moral motivation over rational self-interest. Indeed, the power of this method lies in its ability to capture the seemingly irrational reactions to unfairness observed in the real world (such as when an angry individual spends $10 in fuel costs in order to drive back to a store that he is convinced cheated him out of $5). Interestingly, it seems as if this moral motivation shares neurological real estate with more basic hedonic motivational systems. In studies utilizing measures of brain activation (fMRI) while participants play an ultimatum game, researchers have shown that recipients demonstrate increased activation in reward centers of the brain (including older structures, such as the ventral striatum and amygdala, and the cortical region of the ventromedial prefrontal cortex, which most likely evolved for reasons unrelated to human social cognition) when presented with fair offers as compared to when presented with unfair offers (Tabibnia, Satpute, & Lieberman, 2008). However, those exposed to unfair offers demonstrate increased activation in the bilateral anterior insula, a region commonly associated with the experience of pain and distress (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003). Moreover, the level of activation in brain regions associated with both reward and distress is not predicted by the degree of economic gain or loss resulting from the offer, suggesting that people are sensitive to fairness and unfairness for reasons other than basic economic utility. At the neurobiological level, at least, it is not so much that fairness trumps self-interest, as it is that the brain does little to distinguish between the two.

While research in social and cognitive neuroscience continues to provide evidence for the biological basis of moral evaluations, looking into the brain is only one source of evidence that humans are fundamentally predisposed to evaluate others on a moral dimension. For instance, a great deal of work has demonstrated that humans readily make inferences about the dispositional traits of others given minimal information (Gilbert & Malone, 1995; Gilbert, 1998). Recent work by Todorov and colleagues has shown that this is especially true for inferences regarding trustworthiness. Across a number of studies, these researchers have demonstrated that individuals make quick and automatic evaluations of the trustworthiness of others from brief exposures to their facial features (see Todorov, Said, Engell, & Oosterhof, 2008, for a review). While there is mixed evidence as to whether these evaluations are correlated with actual trustworthiness, a strong bias toward evaluation along this dimension is evidence of its psychological primacy. Moreover, a basic motivation to evaluate others on the dimension of trustworthiness may aid overall accuracy by focusing individuals on other subsequent cues that may be good predictors, such as emotional signals of trustworthiness (Frank, 1988). Such an ability to detect trustworthy individuals would have provided obvious advantages to humanity's ancestors, from allowing them to avoid murderous psychopaths to allowing them to gain social capital by cooperating with trustworthy individuals and avoiding "cheaters."

Fairness and trustworthiness are only a small part of human morality, and it is likely that other facets of morality that have received less empirical attention are just as basic and hardwired. Jon Haidt and his colleagues, for instance, have argued that evolutionary pressures have predisposed us to hold particular moral beliefs regarding in-group loyalty, purity, and authority (Haidt & Joseph, 2004). While cultural influences certainly play an important role in the relative importance placed on these foundational moral intuitions (for instance, politically

conservative individuals are more likely to view violations of in-group loyalty, purity, and authority as moral violations, while liberals tend to focus exclusively on violations of harm and justice), Haidt argues that humans are heavily biased toward perceiving all five of these domains as morally relevant due to humanity's particular evolutionary heritage.

In sum, the mounting evidence that human morality has its origins in biological evolution and is supported by a dedicated set of cognitive mechanisms is consistent with our claim that human moral motivation (especially the motivation to evaluate other individuals on a moral dimension) is "stubborn" and may hold primacy over other kinds of social judgments. It is to the more specific claim that these evaluations appear to trump judgments of freedom that we now turn.

What Kind of Freedom?

We have argued that moral evaluations are a fundamental part of human psychology: humans arrive at moral judgments quite easily, retain confidence in these moral beliefs in the face of challenges to morality, and are strongly motivated to evaluate others on a moral dimension. This moralism is particularly evident in the willingness and ease with which individuals blame and punish others for their moral violations rather than simply ignoring them, even when the cost of punishing outweighs any benefits to the individual (so-called "altruistic punishment"; Fehr & Gächter, 2002). However, people do not go about making these judgments about blame and punishment haphazardly. There is a large body of psychological research describing the underlying rules individuals use in order to determine whether or not to blame others (e.g., Shaver, 1985; Weiner, 1995). These findings can be characterized as confirming an intuition that is most likely shared by most individuals—that when making these judgments,

people care about whether or not an agent acted freely. Specifically, the sort of freedom that seems to matter most for responsibility judgments, and that has been highlighted in psychological theories of responsibility, is what we will refer to as "agentic" freedom (in contrast to the "ultimate," metaphysical freedom argued to exist by libertarian philosophers).

Agentic Freedom. The most influential theories of responsibility within psychology (described briefly above) have generated a great deal of research indicating that when making judgments of responsibility, people seem to care about features of an action that point to an individual's agency (Shaver, 1985; Weiner, 1995; see Vohs, Chapter 5, this volume). These theories emphasize the need for an actor to have possessed volitional control over an action, and require the presence of such things as intentionality, causality, control, and foreknowledge. If all these conditions are met, there is nothing to prevent a judgment that the person should be held responsible and blamed accordingly. These judgments about the local features of an agent and his action, what we are referring to as "agentic freedom," are of obvious importance for at least one simple reason—they allow us to predict the future behavior of an individual with some degree of reliability (something that should matter greatly to a social species living in a group and involved in repeated interactions with the same individuals). Take the criteria of intentionality: if somebody intends and knowingly brings about a harmful outcome, it is a safe bet that she might commit a similar violation in the future, and assigning blame and doling out punishment would serve as an effective social sanction in that it would not only discourage her from committing such violations in the future, but it would also serve as a signal to others that she should not be trusted.

As it turns out, attributing intentionality to actors is something humans do quite readily, and quite often. Intentions seem so central to social judgment that people seem to spontaneously

attribute intentionality even when it is clear that none could exist. For instance, people have been shown to attribute the random movement of shapes on a computer screen to the figures' underlying intentions and desires (Heider & Simmel, 1944), and even infants seem to attribute goal-directed motives to shapes that perform negative and positive behaviors (such as "helping" to push another shape up a hill or preventing the progress of a shape as it tries to climb; Hamlin, Wynn, & Bloom, 2007; Kuhlmeier, Wynn, & Bloom, 2003).

Note that the perception of causality and intentionality in the acts of others most likely did not evolve in the service of attributing moral responsibility but, rather, evolved because it allowed humans to predict the behavior of physical objects and agents they encountered. The detection of causality, for instance, may even be a basic feature of the human visual system. Psychologists studying visual perception have demonstrated that individuals seem to be hardwired to perceive causality and animacy in the movements of simple objects (Scholl & Tremoulet, 2000). Even chimpanzees demonstrate a basic understanding of goal-directed agentic behavior (Premack & Woodruff, 1978), although the evidence that they utilize this understanding in the service of anything other than their immediate self-interest is sparse at best (Call & Tomasello, 2008). While the perception of agency may come easily to us even when presented with the actions of objects and nonhumans, it is nonetheless evident that people are especially likely to perceive agency and intentionality in the acts of other human beings. And if these basic features of agency appear to be completely absent, holding someone morally responsible is much less likely to occur. Take a very simple example: people don't usually blame somebody who brought about harm to another in an entirely accidental way (e.g., if Tom trips on a curb and ends up falling on Dennis, it would seem odd to morally blame Tom for the harm he caused Dennis; at

most someone might accuse him of simple clumsiness and avoid walking too closely to him in the future; Weiner, 1995).

By many measures, psychological theories of responsibility that focus on the requirements of agentic freedom have been successful—they seem to capture a great deal of peoples' intuitions about how and when responsibility should be attributed or attenuated, and do a good job of predicting actual judgments of responsibility across a wide range of cases. As evidence, a great deal of research has demonstrated that when one or more of the criteria that compose agentic freedom are absent, individuals tend to exhibit a reduction in their judgments of responsibility and blame. For instance, relatives of individuals suffering from schizophrenia attenuate blame for actions that were performed as a direct result of uncontrollable hallucinations and delusions (Provencher & Fincham, 2000). Likewise, individuals are more likely to assign blame to AIDS patients if they contracted the disease through controllable means (licentious sexual practices) than if through uncontrollable ones (receiving a tainted blood transfusion; Weiner, 1995). When it comes to the criterion of causality, there is even evidence that people are more sensitive than these theories might predict. For instance, individuals seem to not only care whether or not an agent caused an outcome, but whether he caused it in the specific manner that was intended. If an act was intended and caused, but caused in a manner other than the one intended (so-called "causally deviant" acts; Searle, 1983), participants view the acts as less blameworthy. For instance, a woman who desires to murder her husband by poisoning his favorite dish at a restaurant, but who succeeds in causing his death only because the poison made the dish taste bad, which led to him ordering a new dish to which he was (unbeknownst to all) deathly allergic, does not receive the same blame as if the death were caused by the poison directly (Pizarro, Uhlmann, & Bloom, 2003). It seems as if people are quite capable of paying

close attention to the features of an action in just the manner predicted by traditional accounts of responsibility. In this sense, it is fair to conclude that there is a great deal of evidence that agency fundamentally matters when arriving at judgments of responsibility.

Yet despite the obvious success of these models in predicting a wide range of responsibility judgments, a number of recent findings have emerged that, taken together, point to the conclusion that judgments of agentic freedom are often the result of responsibility judgments rather than their cause. Despite the evidence that humans are capable of making fine-grained judgments about intentions, causality, and control, these distinctions may not actually serve as input when people are strongly motivated to hold another person responsible—the basic mechanisms that give rise to an initial moral evaluation may overshadow the ability or desire to engage in a careful attributional analysis of the sort predicted by these theories. Importantly, the situations that call for making a judgment of responsibility in everyday life may be ones in which people are especially motivated to hold an agent blameworthy, either because of the nature of the violation or because of an initial negative evaluation of an agent.

Even if the psychological processes involved in attributions of agency and those involved in judgments of moral blame are independent, it may still seem odd that judgments of agentic freedom could be distorted to serve the purpose of holding others responsible. However, the criteria used to arrive at a judgment of agency (such as intentionality and control) actually provide a great deal of flexibility and may therefore be especially prone to the influence of motivational biases. In the social domain, the tendency to attribute the cause of others' behavior to underlying intentions often comes at the cost of ignoring causes that are external to the actor (Gilbert & Malone, 1995; Jones & Harris, 1967). The consequence of this attributional bias is that other people's acts are perceived as intentional even if they were not, and any motivation to

blame an individual may exacerbate the perception that an act was intentional. Likewise, the other criteria for agentic freedom, such as possessing control over an outcome, can also be fairly ambiguous—for any given action, there is rarely an easily identifiable, objective answer about how much control an individual truly possessed.

A number of recent findings seem to support the view that judgments of agentic freedom are driven by moral evaluations. Most of this work has focused on the criteria of causality, control, and intentionality and has demonstrated that spontaneous negative evaluations of an act or an agent are enough to change these judgments of agency in a manner consistent with the motivation to blame an agent. For instance, most theories of responsibility posit that possessing causal control over an outcome is an important determinant for the attribution of blame—the less control, the less responsibility (Weiner, 1995). But the relationship between control and responsibility appears far less straightforward. In some cases, individuals impute more control over an act to individuals who seem particularly unlikeable than to other individuals who performed an identical action. Research by Alicke and colleagues, for instance, has shown that individuals make differential judgments about how much control a person had over an outcome if they have reason to think of him as a bad person. In one example, when participants are told that a man was speeding home in a rainstorm and gets in an accident (injuring others), they are more likely to say that he had control over the car if he was speeding home to hide cocaine from his parents than if he was speeding home to hide an anniversary gift despite the fact that the factors that led to the accident were identical across both scenarios (Alicke, 1992, 2000). According to Alicke, the desire to blame the nefarious "cocaine driver" leads individuals to distort the criteria of controllability in a fashion that validates this blame. For Alicke, the

spontaneous and automatic judgments of blame provide a ready source of motivation to distort any information that might be required to justify this blame.

This appears to be true for the intentionality criterion as well. A growing body of research by Knobe and his colleagues has shown that people are more inclined to say that a behavior was performed intentionally when they regard that behavior as morally wrong (Leslie, Knobe, & Cohen, 2006; see Knobe, 2006, for a review). For instance, when given a scenario in which a foreseeable side effect results in a negative outcome, individuals are more likely to say that the side effect was brought about intentionally. In the most common example, the CEO of a company is told that implementing a new policy will have the side effect of either harming or helping the environment. In both cases, the CEO explicitly states that he only cares about increasing profits, not about the incidental side effect of harming or helping the environment (e.g., "I don't care at all about harming the environment. I just want to make as much profit as I can"). Nonetheless, participants perceive that the side effect of harming the environment was intentional—but not the side effect of helping the environment. This pattern of findings (with simpler scenarios) is evident in children as young as 6 and 7 years old (Leslie, Knobe, & Cohen, 2006). Although the mechanism underlying these findings that morally bad actions are perceived as more intentional have been hotly debated, the findings are consistent with the view we are defending here—that the motivation arising from a desire to blame leads people to make judgments of freedom that are consistent with this blame.

If the motivation arising from a judgment of moral badness leads to a magnification of intentionality judgments, then it should be the case that individuals who find a certain act particularly bad should judge these acts as more intentional than individuals who are indifferent about the morality of the act. In a series of studies, Tannenbaum, Ditto, and Pizarro (2009)

demonstrated that this is indeed the case. Across a number of studies, individuals who differed in their initial assessment of how immoral an act was demonstrated differing judgments of intentionality. For instance, using the scenario developed by Knobe and colleagues described above, individuals who reported a strong moral motivation to protect the environment were more likely to report that the damage to the environment brought about as a side effect of the CEO's decision was intentional than those who held no such moral values about the environment. Similarly, when presented with a case in which military leaders ordered an attack on an enemy that would have the foreseen (yet unintended) consequence of causing the death of innocent civilians, political conservatives and liberals rated the action as intentional when it ran contrary to their politics. When the case involved Iraqi insurgents attacking American troops with the side effect of American civilian deaths, political conservatives were more likely to judge the killings as intentional than liberals. The converse was true when the case was described as American troops killing Iraqi civilians as a consequence of attacking Iraqi insurgents.

In another study, participants were given one of two examples in which the principle of harm reduction was used to justify the distribution of free condoms in order to prevent the incidence of pregnancy and the spread of disease. Participants were all subsequently asked whether the distribution of condoms intentionally promoted sexual behavior. Moral motivation was manipulated by changing the intended recipients of the free condoms: one set of participants read about local school board members who decided to distribute condoms to students because they were aware that middle-school and high-school students were engaging in risky sexual behavior, while another set read that policy makers who were aware that soldiers stationed in foreign countries were engaging in acts of sexual aggression (i.e., rape) against local women decided to distribute condoms to the soldiers in order to curb the spread of unwanted diseases. In

both cases it was explicitly stated that the group of individuals making the decision to distribute condoms did not condone the sexual behavior of the intended recipients but merely wanted to reduce the harm associated with these actions.

As predicted, participants who read about the distribution of condoms to young teens reported that the policy makers were not intentionally promoting teen sex, while participants reading about foreign soldiers reported that policy makers had intentionally promoted rape. The obvious difference in the moral status of premarital teen sex and rape seemed to be driving judgments of intentionality. More interestingly, these differences in the moral status of the two acts may have accounted for an unexpected gender difference in these ratings of intentionality—women were much more likely to report that the policy makers had intentionally promoted rape in the military case but had not intentionally promoted sex in the case of young teens. Men judged the two scenarios nearly identically. One interpretation of this finding is that the motivation to condemn rape was stronger in female participants, especially since the sexual aggression in the scenario targeted women, and that this increased motivation drove intentionality judgments.

Consistent with the view that a motivation to hold an individual blameworthy can lead to a distortion of the criteria for agentic freedom, it appears as if blame can even influence memory for the severity of a moral infraction. In one study, Pizarro and colleagues presented individuals with scenarios in which a man walked out of a restaurant without paying for his meal. Within the description of this action, participants were given detailed information about the price of the dinner. Researchers manipulated the degree of blameworthiness for this infraction by telling one set of participants that the individual had failed to pay for his meal because he had received a call notifying him that his daughter had been in a car accident, while another set of participants were

told that the individual simply desired to get away without having to pay. When asked to recall the events approximately one week later, participants who had read that the man failed to pay simply to get away with it recalled the price of the dinner as significantly higher than those who had read that the man had left because of his daughter's accident (whose memory for the price was accurate). Across conditions, the degree of moral blame participants reported after reading the story was a significant predictor of the memory distortion one week later (Pizarro, Laney, Morris, & Loftus, 2006).

"Ultimate" Freedom. The agentic freedom that seems important for lay attributions of moral responsibility is conceptually independent from the "ultimate," metaphysical freedom that many philosophers have argued is necessary for attributing moral responsibility (e.g., libertarian freedom, or the freedom to "have done otherwise"). A deep concern espoused by many is that the causal determinism that has been a reliable guide in scientific research may also threaten the ultimate freedom that seems necessary to be held morally accountable. If thoughts and feelings are entirely caused by physical processes in the brain, and if the laws that govern these processes are no different than the laws governing the motion of billiard balls and automobiles, then perhaps a person is no more accurate in the belief that she freely decided to get up in the morning than that a billiard ball freely chose to roll across the table. Some have argued that the increased dissemination of psychological research highlighting the deterministic processes that give rise to human thought and action may radically change people's notions of freedom and punishment (Greene & Cohen, 2004). Yet while psychologists have conducted a great deal of research on the criteria of agentic freedom, only recently have psychologists and experimental philosophers turned their attention to the question of whether this ultimate freedom is treated by individuals as

a prerequisite for the ascription of moral responsibility in the manner many philosophers have argued.

As it turns out, the deep concerns that determinism threatens moral responsibility may be unfounded. In the first place, it is unclear whether the lack of ultimate freedom is of much concern for individuals when making judgments of responsibility—in some cases, despite explicit information that undermines the presence of ultimate freedom, people still seem willing to hold others morally accountable. Yet even in cases in which participants appear to endorse the view that ultimate freedom is a necessary prerequisite for the attribution of moral responsibility, the motivation to hold others morally accountable seems to lead individuals to selectively ignore this information. For instance, in a recent study by Nichols and Knobe (2007), individuals were presented with a description of a world that was described as entirely deterministic in the manner that is incompatible with ultimate, metaphysical freedom. When participants were asked if, in general, murderers in this deterministic world should be held morally responsible, most participants said no. But when presented with a specific individual who murdered his entire family, individuals were more than willing to attribute blame—even when it was clear from the description of the world that he could not have acted otherwise. It appears as if the vivid, emotional nature of the specific crime (and the likely motivation to hold an individual capable of such a crime given its vividness) led individuals to either ignore the information about ultimate freedom or adjust their metaphysical beliefs about whether determinism is truly incompatible with moral responsibility.

In support of this latter view, Woolfolk, Doris, and Darley (2006) described a scenario to participants in which a man was under a clear situational constraint that forced him to murder a passenger on an airplane (he was forced by hijackers to kill the person or else he and 10 others

would be killed). While the man appeared to possess agentic freedom, he was clearly constrained by the situation in terms of his ability to have acted otherwise. Despite this constraint, participants nonetheless held the man more responsible for the murder if it was something he wanted to do anyway (if he "identified" with the act). Consistent with the compatibilist approach espoused by Frankfurt (1971) and others, the man's inability to do otherwise—his lack of ultimate freedom—did not seem to disturb participants' ability to hold him morally accountable. As a final example of how underlying motivation to hold an individual accountable can shift judgments of freedom, we recently asked participants to recount a bad deed that was either committed by a good friend or by an enemy (Helzer & Pizarro, 2009). Given the differing motivations at play in their judgments of responsibility for a liked versus a disliked individual, we were interested in how individuals would judge the freedom of the person who had committed the moral infraction by asking participants to judge to what extent they thought the person's behavior was intentional as well as whether they thought the person's behavior was freely chosen. As expected, people found the bad deed performed by an enemy as more blameworthy than the one performed by a friend (this was true after controlling for the severity of the act, as well as when limiting our analyses to judgments of identical acts committed by a friend or an enemy, such as sexual infidelity). More importantly, the manipulation of motivation (friend vs. enemy) affected participants' judgments of freedom as assessed by their responses to both questions: relative to the immoral deeds of their friends, participants attributed the misdeeds of their enemies to their enemy's underlying intentions and were more likely to report that the act was freely chosen.

What these studies seem to show is that while many philosophers perceive determinism to be a threat to freedom (and by extension, to moral responsibility), the psychology of moral

responsibility is such that people are fairly willing to hold others morally accountable even in the presence of strong determinism (see also Nahmias, Morris, Nadelhoffer, & Turner, 2006, for evidence suggesting that for laypersons, determinism doesn't even threaten freedom of the libertarian variety).

Conclusion

We have tried to defend a number of specific claims about the lay concept of freedom and how individuals use information about freedom when making judgments of moral responsibility. Specifically, we have argued that individuals are highly motivated to hold others accountable for moral infractions, and that the primacy of this motivation often influences judgments about freedom, rather than the other way around. This seems true for judgments regarding agentic freedom (such as whether an act was intended, caused, and controlled), as well as for judgments of ultimate, or metaphysical freedom (the ability to have done otherwise). One upshot of these findings is that the classic incompatibilist view that determinism poses a threat to morality and responsibility, while seemingly intuitive, may pose a threat only to individuals who do things like read books about free will. The psychological link between ultimate freedom and responsibility appears less strong than many have suggested. Indeed, to the extent that individuals possess the intuition that there is a link between freedom and responsibility, they seem to use such a link primarily as a strategy to defend their moral judgments. Free will is important, but not for the reason many might think.

In moments of reflection, then, individuals may realize that they should suspend their moral evaluations of the robot neighbor described in the introduction (because he does not possess "ultimate" freedom). Nonetheless, when he plays his music loudly, fails to mow his

lawn, and lets his dog use their yard as a bathroom, they will have no problem attributing all the freedom needed to in order to blame him for his rude behavior.

Discussion with David A. Pizarro

Are there differences between beliefs about free will in the abstract and beliefs about free will in concrete cases?

Bertrand Malle conducted several studies in which he asked people abstract questions about people's intuitions about intentionality, and results seemed to indicate that people believe that a person has to intend to do a thing in order to do it intentionally. In contrast, Joshua Knobe gave participants a concrete scenario featuring a CEO and asked them to make determinations about whether his actions were done intentionally. Results indicated that one need not intend to do something in order to do it intentionally. Thus, there seems to be a distinction between abstract beliefs and the concrete case for the question of intentionality. Might there be a similar distinction for free will?

Indeed, people do seem to respond differently to questions about free will in the abstract as compared to concrete cases. In particular, people generally believe both that determinism undermines free will in the abstract. Yet, for specific cases, people generally believe that individuals are not to be excused from wrongdoing because of determinism. This seeming conflict likely is because people have a strong urge to see intentionality, in order to hold others responsible. Said differently, people seem to alter what they mean by intentionality, or the requirement of free will in establishing intentionality, in order to blame. It would be unusual to hear a judge say "Because of metaphysical determinism, you should get a reduced sentence." That is, abstract beliefs about free will may conflict with the desire to assign blame in concrete cases.

Is the question of free will related to judgments of moral responsibility?

The consensus view is that the two are deeply entwined. However, an alternative view is that free will and moral responsibility are entirely orthogonal, such that determined bad behavior requires punishment just as undetermined bad behavior does. For instance, if a parent has one child who is temperamentally (and unavoidably) agreeable, and one that is temperamentally (and unavoidably) disagreeable, the parent will surely increase the punishment on the disagreeable child, despite the fact that it is temperamental and not chosen. Indeed, greater punishment may be necessary to alter the disagreeable child's behavior. On this view, there is no real relationship between freedom and responsibility.

A second alternative to the view that free will and moral responsibility are closely related is that they are peripherally related. On this view, the question of free will is so much more important than that of moral responsibility that moral responsibility is incidental. Consider the following example. Imagine a person who is totally amoral. This person makes no moral judgments about others or about the self. Such a person may experience interpersonal problems, but he or she can function in society and order from a menu. In contrast, one cannot function without a sense of free will. A person convinced of the truth of determinism cannot go to a restaurant, sit down and say "Everything is determined, so I'll just wait and see what I order". In other words, people cannot function without the assumption of free will, but they can function without the concept of moral responsibility. Thus, moral responsibility may be considered a peripheral question to that of free will.

Laypersons, however, seem to react strongly and viscerally to the possibility of an absence of moral responsibility, while they view the possibility of determinism with less

reactance. Thus, among laypersons, the question of moral responsibility seems preeminent relative to the question of free will.

Is moral responsibility unique among humans?

One of the distinguishing features of human beings is self-regulation. In most nonhuman animals, behavior is under social control. However, certain patterns of instinctual behavior are adaptively necessary for the survival of the species. The way that individual nonhuman animals are controlled is by how their behavior affects the group. This is social control. The interesting thing about judicial decisions is that they are partly a system of social control and partly a system that has tapped into the notion of self-regulation and moral responsibility. In a sense those are confounded roles. The distinction between the kind of social control found in nonhuman animals and the kind of social control found in humans is that an additional amount of social control is found in humans.

References

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology, 63*, 368–378.

Alicke, M. D. (2000). Culpable control and the psychology of blame. *Psychological Bulletin, 126*, 556–574.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.

Baumeister, R. F., Masicampo, E. J., & DeWall, C. N. (2009). Prosocial benefits of feeling free: Disbelief in free will increases aggression and reduces helpfulness. *Personality and Social Psychology Bulletin, 35,* 260-268.

Brosnan, S. F., & deWaal, F. B. M. (2003). Monkeys reject unfair pay. *Nature, 425*, 297-299.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Science, 12*, 187-192.

Clarke, R., (2003). *Libertarian accounts of free will*. New York: Oxford University Press.

DeWaal, F. (2006). *Primates and philosophers. How morality evolved.* Princeton, NJ: Princeton University Press.

Dennett, D. (1984). *Elbow room: The varieties of free will worth wanting.* Cambridge, MA: MIT Press.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*, 137–140.

Frank, R. (1988). *Passions within reason: The strategic role of the emotions.* New York: W.W. Norton.

Frankfurt, H. (1971). Freedom of the will and the concept of the person. *Journal of Philosophy, 68,* 5–20.

Gilbert, D. T. (1998). Ordinary personology. In D. T. Gilbert, S. T., Fiske, & G. Lindzey (Eds.) *The handbook of social psychology* (4th edition). New York: McGraw Hill.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21–38.

Greene, J. D., & Cohen J. D. (2004). For the law, neuroscience changes nothing and everything. *Philosophical Transactions of the Royal Society of London B (Special Issue on Law and the Brain), 359*, 1775–17785.

Haidt, J., & Joseph, C. (2004). Intuitive ethics: How innately prepared intuitions generate culturally variable virtues. *Daedalus*: *Special Issue on Human Nature,* 55–66.

Haidt, J., Rosenberg, E., & Hom, H. (2003). Differentiating diversities: Moral diversity is not like other kinds. *Journal of Applied Social Psychology*, 33, 1–36.

Hamlin, J. K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants. *Nature, 450,* 557–559

Heider, F., & Simmel, S. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57*, 243–259.

Helzer, E. G., & Pizarro, D. A. (2009). *Motivated Use of Ultimate versus Agentic Freedom*. Manuscript in progress.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24.

Knobe, J. (2006). The concept of intentional action: A case study in the uses of folk psychology. *Philosophical Studies, 130*, 203–231.

Kuhlmeier, V., Wynn, K., & Bloom, P. (2003). Attribution of dispositional states by 12-month-olds. *Psychological Science, 14*, 402–408.

Leslie, A. M., Knobe, J., & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science, 17*, 421–427.

Miller, G. F. (2007). Sexual selection for moral virtues. *Quarterly Review of Biology, 82*, 97–125.

Nahmias, E., Morris, S., Nadelhoffer, T., & Turner, J. (2006). Is incompatibilism intuitive? *Philosophy and Phenomenological Research, 73*, 28–53.

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

Pereboom, D. (2001). *Living without free will*. New York: Cambridge University Press.

Pizarro, D. A., Inbar, Y., & Darley, J. M. (2009). *A dual-process account of judgments of free will and responsibility*. Manuscript in preparation.

Pizarro, D.A., Laney, C., Morris, E., & Loftus, E. (2006). Ripple effects in memory: Judgments of moral blame can distort memory for events. *Memory and Cognition, 34*, 550-555.

Pizarro, D.A., Uhlmann, E., & Bloom, P. (2003). Causal deviance and the attribution of moral responsibility. *Journal of Experimental Social Psychology, 39*, 653-660.

Premack, D., & Woodruff, G. (1978). Does the Chimpanzee Have a Theory of Mind. *Behavioral and Brain Sciences, 1*, 515-526.

Provencher, H., & Fincham, F. D. (2000). Attributions of causality, responsibility, and blame for positive and negative symptom behaviors in caregivers of persons with schizophrenia. *Psychological Medicine, 30,* 899–910

Roskies, A. L. (2006). Neuroscientific challenges to free will and responsibility. *Trends in Cognitive Sciences*, *10*, 419–423.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the Ultimatum Game. *Science, 300*, 1755–1758.

Scholl, B. J., & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Sciences, 4*, 299–309

Shaver, K. G. (1985). *The attribution of blame: Causality, responsibility, and blameworthiness*. New York: Springer-Verlag.

Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind.* New York: Cambridge Press.

Strawson, P. F. (1974). *Freedom and resentment and other essays*. London: Menthuen Publishing Ltd.

Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding fairness activates self-control circuitry. *Psychological Science, 19*, 339–347.

Tannenbaum, D., Ditto, P.H., & Pizarro, D.A. (2009). *Motivated assessments of intentionality.* Manuscript in progress.

Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences, 12*, 455-460.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, *46*, 35–57.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science, 19*, 49-54.

Weiner, B. (1995). *Judgments of responsibility: A foundation for a theory of social conduct*. New York: Guilford Press.

Woolfolk, R. L, Doris, J. M., and Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition, 100*, 283–301.