Bringing Character Back:

How the Motivation to Evaluate Character Influences Judgments of Moral Blame

David A. Pizarro
Cornell University

David Tannenbaum
University of California, Irvine

Human beings are deeply moral creatures. Perhaps nowhere is this more evident than in the stories we tell. Literature, cinema, and television are replete with tales that (either literally or metaphorically) describe the battle between good and evil, and tell the stories of the heroes and villains fighting for each side. But while we may root for the heroes, it is the villains that often capture most of our attention. Like audiences in the silent-movie era, who would boo and hiss loudly when the villain appeared onscreen, not only are we motivated to condemn the villain for his immoral actions, we seem to take great pleasure in doing so. For those of us of a certain age, there was one villain who allowed us this pleasure more than any other: Darth Vader, the antagonist of the original "Star Wars" films. From the moment he stepped onto the screen, there could be no doubt in the audience's mind that he was the bad guy. Vader exuded all of the cues used by moviemakers to communicate "evil": He was clad entirely in black, spoke with a deep, ominous voice, and was as much a machine as a human being. To be sure, if we ever encountered him in real life we would have been very motivated to keep a safe distance.

The motivation to identify and condemn villains is not limited to our role as audience members, however. Few tasks are as important to our social well-being as figuring out who the

"good guys" and the "bad guys" are in our everyday lives. Many social decisions require us to make an evaluation regarding a person's underlying traits — such as trustworthiness, honesty, compassion, or hostility — that together constitute an individual's moral character. For instance, how do I know whether or not to trust the person trying to sell me a car? Should I accept a date with someone I've just met, or is he or she a creep? Should I believe the teenager at my door who says his car broke down and he just needs to use my phone? Getting these evaluations right is important; misreading a person's character might not only lead to poor financial or romantic decisions, it could get one killed. And these judgments are not just one-shot deals. Keeping track of the good people and the bad people are over time is just as important, lest we get cheated again by the same person or unwittingly offer help to someone who might never help in return.

Unfortunately, unlike Darth Vader the bad characters we encounter in everyday life do not always dress in black or speak in ominous voices, so figuring out whether someone possesses negative character traits is more complicated than spotting a cinematic villain (who are more akin to caricatures of "pure evil,"; Baumeister, this volume). Yet despite the lack of such overt cues, we seem motivated and well-equipped to evaluate other people's underlying character traits. In fact, there is a great deal of evidence that these evaluations are psychologically primary. We evaluate agents on the dimension of goodness or badness automatically and with little effort starting remarkably early in life, and this seems to be true of individuals across cultures (Bloom, this volume; Fiske, Cuddy, & Glick, 2007; Hamlin, Wynn, & Bloom, 2007; Kuhlmeier, Wynn, & Bloom, 2003; Willis & Todorov, 2006). Before we shake a person's hand for the first time, we have most likely already made a judgment about his or her trustworthiness (Todorov, Said,

Engell, & Oosterhof, 2008), and noticed whether he or she appears hostile or threatening (Bar, Neta, & Linz, 2006).

Moreover, we continue evaluating others' character long after our first encounter through a variety of methods, such as observing their emotional signals (Ames & Johar, 2009; Frank, 1988) or gossiping with friends about the others' moral failings (Foster, 1997). This desire to track others' character is also evident in our concern for people's reputations. Reputations affect our ability to succeed in games based on trust and cooperation (e.g., Rand, Dreber, Ellingsen, Fudenberg & Nowak, 2009). The motivation to keep track of the other's character is evident even in memory: We have better memory for the faces of people who cheated us unexpectedly (or helped us unexpectedly; Chang & Sanfey, 2009).

This ability and motivation to evaluate others on the basis of moral character was likely of such fundamental importance during primate and human evolution that it is most likely a product of natural selection. For instance, to the extent that moral character was predictive of whether a person would cooperate or defect in joint endeavors, character assessment was invaluable when making social decisions that directly affected survival and reproduction (Gintis, Henrich, Bowles, Boyd, & Fehr, 2008). More generally, individuals who were able to detect the presence of underlying moral traits in others would have been better able to avoid cheaters, psychopaths, and murderers, and would have benefited from forming reliable social relationships with trustworthy individuals who could provide help when needed. Recently, Miller (2007) argued that sexual selection pressures may have favored the ability to evaluate the character traits of potential mates. To the extent that these traits were correlated with future choices, such as

parental investment, the survival and reproduction chances of one's offspring might depend on valid character assessments during mate selection.

In short, the motivation to evaluate others' character appears to be a fundamental feature of human social cognition, and for good reasons. Accordingly, one would expect that theories of moral judgment — particularly those that focus on how we evaluate the others' moral actions — might place great emphasis on how such character evaluations influence moral judgment. Yet this is not the case. In what follows, we argue that theories of moral judgment (specifically, theories of moral blame) are fundamentally incomplete because they disregard the primacy of character evaluations. We then outline an alternative character-based theory of moral blame that may explain recent findings in the literature on moral responsibility that were incorrectly viewed, in terms of previous theories, as judgment errors, but appear natural in light of motivation to evaluate other's moral character. By integrating character into the psychology of moral judgment, we hope to arrive at a more accurate account of *how* we make judgments of moral blame by taking into account *why* we make these judgments in the first place.

The Psychology of Moral Blame

A quick read of a daily newspaper, or a few minutes eavesdropping at the office water cooler, is probably sufficient to convince anyone that moral judgments come quickly and easily to most human beings. Yet figuring out how we make these judgments has proven difficult. One longstanding puzzle in the study of morality has been the wide variety of beliefs among individuals and across cultures concerning which acts are immoral. Why, for instance, do some people believe that aborting a fetus, torturing a prisoner for information, or pirating music are moral "don'ts," while others not only disagree with these beliefs but go to great lengths to defend

their opposites? Answering this question — how and why we come to believe that certain acts are morally taboo, permissible, or obligatory — has been a central concern of many moral psychologists (Cushman & Greene, this volume; Graham & Haidt, this volume; Greene, 2003; Haidt, 2001; Inbar, Pizarro, Knobe & Bloom, 2009). But even when an act is uncontroversially perceived to be morally wrong, we often have to make an additional moral judgment to determine whether the person who has committed the act should be held morally responsible. These two judgments — moral acceptability and responsibility—are the basis for judgments of blame (blame being an ascription of responsibility for a morally bad action). A number of highly influential theoretical accounts have been proposed to describe and explain how such judgments are made (Shaver, 1985; Weiner, 1995). It is to these accounts of blame that we now turn.

Moral Blame: The Standard Account

Most moral infractions we encounter in everyday life are minor: Someone cuts in front of you while in line at the grocery store, unfairly insults a sensitive co-worker, or spreads questionable rumors about a friend. But we are also confronted not infrequently with more serious infractions, even if only when watching the nightly news or reading the newspaper (e.g., a mother drowns her child or a man is convicted of embezzling company funds). Although we generally believe that cutting in line is wrong and killing a child is very wrong, we do not always hold people responsible for such acts. Maybe the person cut in front of you failed to see you; maybe the mother who drowned her child was the victim of a mental illness. It is important to get these judgments right, because judgments of right and wrong carry social sanctions such as exclusion, imprisonment, and in some cases even death.

The normative answer to this question of how blame should be assigned—that is, how we *ought* to make these judgments — has been discussed by philosophers and legal scholars for centuries (e.g., Aristotle, 4th Century B.C.E.; Hart, 1968). For psychologists, these normative theories have served as a starting point for developing more complete theories of responsibility and blame. For instance, the earliest and most influential theories of moral psychology, the developmental theories of Piaget and Kohlberg, were heavily influenced by Kant's (1796/1996) deontological ethics, according to which the moral status of an act is evaluated in relation to rules, duties, or obligations viewed as a set of constraints on action (Ditto & Liu, this volume; Kagan, 1997). In a deontological approach, actions are viewed as morally impermissible if they violate these constraints (such as the prohibition against knowingly taking an innocent life). This view implies that to be held responsible for an act, an individual must have had the ability to do otherwise. In the absence of the freedom to act differently, holding an individual blameworthy would be unjustified. In Kant's view, *ought* implies *can*. If an individual had no control over an action, or did not intend or foresee the infraction, he or she could not have acted otherwise and is therefore not blameworthy (Bayles, 1982).

The deontological approach has been contrasted with the equally influential consequentialist approach to ethics (e.g., Ditto & Liu, this volume; Smart & Wiliams, 1973), which makes no distinctions regarding rules, duties, and obligations but proposes one criterion for evaluating the moral "rightness" or "wrongness" of an act — whether or not it brings about a favorable outcome. Moral acts, then, are defined as ones that maximize "good" consequences, and ones that avoid negative consequences. One upshot of this view is that moral blame is relevant only insofar as it might socially sanction and deter future negative acts. For the

consequentialist, it is permissible for sanctions to be imposed whether or not the offender could have done otherwise. Features important from a deontological perspective then, such as the specifics of an individual's mental state, are in-and-of themselves meaningless for determining sanctions.

Of importance to the argument we are advancing here, both of these normative approaches place very little (if any) emphasis on evaluations of a person — they are fundamentally *act-based* rather than *person-based*. They propose that moral evaluations should focus on local features of an act and agent (e.g., whether the action violates a rule, whether the agent's mental state at the time of the action allowed for alternative actions, or whether the act caused harm). In contrast, a person-based approach would take the person as the unit of analysis when judging blame — their underlying traits, dispositions, and character (Bayles, 1982). Such an approach seems to fit our normal (and biologically ancient) reasons for blaming, because it takes into account the goal of removing "bad people" from important positions in our social lives. And there is a theory in normative ethics that takes this view: virtue ethics (e.g., Anscombe, 1958). This approach emphasizes the character of the agent, rather than whether an act complies with rules or has good consequences. In fact, the claim that morality is fundamentally about possessing the right kind of character can be traced (at least in Western thought) to the views of Plato and Aristotle, who argued that to be a moral person means to have a moral character, or to possess desired virtues. Although this view fell out of favor among philosophers as the deontological and consequentialist approaches gained ground, the virtue-based approach has enjoyed a resurgence in philosophy, and in legal theory as well (this resurgence has been referred to as the "Aretaic Turn"; Solum, 2004). So far, however, this virtue-

based approach to ethics has gained little ground in moral psychology (but see Monin, Pizarro, & Beer, 2007).

By building on deontological and consequentialist normative approaches, psychological theories of blame have inherited their act-based approach to moral assessment, inasmuch as they outline a set of local criteria for determining responsibility for a moral infraction (and hence blame; Shaver, 1985; Weiner, 1995). In addition, consistent with the attribution theories from which they emerged (e.g., the theories of Heider, 1958, and Kelly, 1967), psychological theories of blame have assumed that, when given the necessary information, lay judges are capable of determining whether or not these criteria were met in any given act. That is, when presented with an instance in which a moral infraction was committed, the lay judge is presumed to work his or her way through the criteria in a stage-like fashion, asking a series of questions about features of the act, such as whether the actor intended the outcome, had control over the outcome, or could foresee the results of the action. If these conditions are met, there is nothing to prevent a confident judgment that the person should be held responsible and blamed (or praised, in the case of positive actions) accordingly. However, if some of these criteria are not met (e.g., the agent did not intend the outcome), these theories predict that the lay judge will either attenuate blame or ascribe no blame at all. It should be noted that these theories assume an invariant application of these decision rules across similar judgments; the same criteria should be applied regardless of time, place, or individual (Ditto & Liu, this volume; Knobe & Doris, in press). For instance, when determining whether an individual should be blamed for stealing a car, his capacity to distinguish right from wrong and his ability to form intentions matter, but whether he is the judge's best friend or worst enemy should not matter. Likewise, if a person accidentally trips and

knocks another in the face with her arm, whether or not she has a criminal record bears little on the assessment of blame because she had little control over the outcome.

The criteria outlined by these theories of blame seem intuitively reasonable, and the theories have fared quite well in predicting judgments of responsibility across a wide range of cases. When one or more of the designated criteria for blame are absent in a given case, research participants tend to reduce the amount of blame the assign to the agent. For instance, relatives of individuals suffering from schizophrenia reduce the blame they assign for harmful actions undertaken as a result of the individual's (uncontrollable) hallucinations and delusions (Provencher & Fincham, 2000). And research participants are more likely to assign blame to AIDS patients if they contracted the disease through controllable means (licentious sexual practices) than if they contracted it uncontrollably (receiving a tainted blood transfusion; Weiner, 1995). In addition, unintentional acts, such as accidental harms, are seen as less blameworthy than intentional acts, and acts that are unforeseeable as less blameworthy than foreseeable acts (Weiner, 1995).

When it comes to the issue of causality, people are more sensitive than even these classic theories might have predicted. For instance, individuals seem not only to care whether an agent caused an outcome, but whether he caused it in the specific manner in which he intended. If an act was intended and caused, but caused in a manner other than the one intended (acts that have been referred to as "causally deviant"; Searle, 1983), research participants view the acts as less blameworthy. In one study, for example, Pizarro and colleagues (Pizarro, Uhlmann, & Bloom, 2003) presented participants with the story of a woman who desired to murder her husband by poisoning his favorite dish at a restaurant, but she succeeded in causing his death only because

the poison made the dish taste bad, which led to him to order a new dish to which he was (unbeknownst to all) deathly allergic. In cases like these, participants did not assign the same degree of blame as if the outcome had been caused directly in the manner the agent had intended. It seems, in short, that people often pay very close attention to the features of an action in just the manner described by deontological and consequentialist theories of blame.

A Character-Based Alternative to Understanding Blame

Despite the empirical support these theories have received, a number of recent findings have called their accuracy into question. For instance, judgments of moral blame are often disproportionate to the actual harm an agent caused; relatively harmless acts can receive harsh moral judgments. In addition, the mental-state criteria used to determine blame do not always fit the stage-like pattern predicted by the traditional approaches. In fact, research participants' judgments are often influenced by information that the traditional theories consider extraneous and irrelevant, such as the outcome of the act or the characteristics of the person performing the act (e.g., Alicke, 2000; Knobe, 2006).

One way to interpret these findings is to take them as evidence that, as in other judgmental domains, people are prone to error and bias in their judgments of moral blame. For example, rather than carefully taking the proper criteria into account before making a judgment of blame, people are affected by the emotions aroused by certain acts (e.g., Alicke, 2000). We believe, however, that such findings represent more than just a growing catalog of "errors" in moral judgment — simple deviations from otherwise accurate theories of blame. We believe that there are systematic patterns in the "errors" suggesting that the theoretical approaches themselves are error-prone rather than the people making the judgments. This is why we are proposing a

person-based character approach as an alternative to the act-based theories. This approach can explain putative judgmental "errors" as the systematic (and often rational) output of a system that is primarily concerned with evaluating others' character traits. A simple way of highlighting the difference is to wonder whether the person making a judgment of blame is asking him- or herself "Was this particular action wrong?" or "Is the person who committed this act a bad person?"

We want to argue that the motivation to evaluate an agent's character manifests itself in at least two related ways when one is presented with a moral infraction. First, to the extent that a given act seems diagnostic of negative character traits, the agent of the act is more likely to be seen as deserving of blame. This may lead to harsh judgments for actions that seem fairly harmless in themselves but are indicative of a "bad" character. Second, if there is information about an individual's character that is extrinsic to the features of a particular act, it will be applied in judgments of blame (including judgments of such issues as control, causality, and intentionality). For instance, if there is evidence that an individual is a bad person, the inference that he or she intended a negative outcome seems reasonable (because bad people, by definition, are likely to desire and intend bad things).

*Asymmetries in judgments of control, intentionality, and blame***.** Extant theories of blame make a straightforward prediction that criteria such as control, intention, and causality feed directly into judgments of blame. If, for example, an individual has absolutely no control over an action (and simply could not have done otherwise), he or she should not be held responsible. If caffeine jitters cause you to accidentally donate money to charity by clicking on the wrong computer key, or you accidentally scratch your friend's car with your key because someone bumped into you, you are not a candidate for praise or blame.

A number of studies have indicated that the relation between these criteria and judgments of blame are not so simple. Despite evidence that humans are capable of making fairly careful distinctions regarding the presence of intentions, causality, and control, these distinctions may be overshadowed by a negative evaluation of character. This evaluation may cause "inflated" judgments of intentionality, causality, and control in cases where an agent seems particularly nefarious. Given the argument that these good/bad character judgments are psychologically primary (perhaps for evolutionary reasons), this should come as little surprise, but this asymmetric ascription of increased intentionality, causality, and control is puzzling given a standard act-based account of moral judgments.

Research by Alicke and colleagues (see Alicke, 1992, 2001) has shown that we make differential judgments about how much control a person had over an outcome if we have reason to think of him as a bad person. In one study, participants were told that a man was speeding home in a rainstorm and got into an accident that injured others. When asked whether the accident was due to factors under the driver's control (e.g., he was driving irresponsibly), participants were more likely to agree if they were previously told that he was speeding home to hide cocaine from his parents than if they were told he was speeding home to hide an anniversary gift, despite being given identical information regarding that the factors that led to the accident. According to Alicke, our desire to blame the nefarious "cocaine driver" is what leads us to distort the criteria of controllability to validate this blame. Again, on the standard act-based view of responsibility, this appears to be a bias in judgment. But on the character based-account, this makes sense. If we have just been provided with information that an individual is the sort of person to be hiding cocaine in his parents' house, it seems reasonable to assume that he might be

the sort of person who drives recklessly. That is, given minimal, incomplete, or ambiguous information about controllability or intentionality, we are likely to take character information into account when asked to arrive at an estimate of these features. In fact, because we rarely have a window into factors such as the mental state of the individual at the time of the infraction (there is often not an easily identifiable, objective answer to the question of how much control an individual actually possessed), it seems as if applying information about an individual's previous acts, his or her known behavioral tendencies, or his or her character traits is a valid (albeit not perfect) way to make an assessment, much as we would apply base rate information when making other kinds of judgments under uncertainty.

This filling-in-the-gaps using information about an individual's character may also be at work in intentionality judgments. Research by Knobe and his colleagues (Leslie, Knobe, & Cohen, 2006; see Knobe, 2006, for a review) has shown that people are more likely to say that an act was performed intentionally if they perceive it to be morally wrong. In many of Knobe and colleagues' examples, individuals were provided with a scenario in which a foreseeable side-effect results in a negative outcome. They were then asked if the side-effect was brought about intentionally. For instance, in one scenario participants were told that the CEO of a company decided to implement a new policy, but that the policy would have the side-effect of either harming or helping the environment. Across both versions of the scenario (harm the environment or help the environment), participants were told that the CEO explicitly said cares only about increasing profits, not about the incidental side-effect of harming or helping the environment ("I don't care at all about harming the environment. I just want to make as much profit as I can"). Nonetheless, participants judged the side-effect of harming the environment as intentional, but

not the side-effect of helping. This pattern of findings (with simpler scenarios) is evident in studies involving children as young as six or seven years old (Leslie, Knobe, & Cohen, 2006).

From an act-based approach this so-called "side-effect effect" is puzzling, because judgments of intentionality are thought to be descriptive judgments about an agent's state of mind, and they should therefore be independent of the moral implications associated with the act. For these reasons, several researchers (including one of the authors: Pizarro & Helzer, in press) have concluded that the side-effect effect is the result of a bias or "performance error" in the way our intentionality judgments are made. However, other researchers (including Knobe) have suggested that intentionality judgments may be more than just assessments of mental states, and instead may be fundamentally imbued with normative considerations of praise and blame.

In a similar vein, Wellman and Miller (2008) have argued that deontic considerations (judgments of permissibility or obligation) are fundamental to reasoning about intentionality (or belief-desire reasoning, which is usually thought to be a core component of intentionality; see Malle & Knobe, 1997). That is, obligations regarding harming and helping—obligations held to be especially important for morality—are asymmetric, such that we perceive a greater duty not to cause harm than we do to help (Grueneich, 1982). It makes sense, then, that we perceive acts in which an agent foresees a potential harm as different from those in which the agent foresees a potential benefit. The CEO in the "harm" side-effect example above is performing a behavior that runs counter to the strong obligation to avoid knowingly causing harm. Because he continues to carry out his chosen action despite this obligation, it is reasonable to infer that his behavior was performed intentionally (overriding a strong obligation seems to require greater intentionality than overriding a weak one). However, because the CEO in the "help" condition

does not have a strong obligation to help, it is reasonable to infer that his behavior is less intentional.

What this means from a character-based approach is that people will judge cases in which a side-effect causes harm as being particularly diagnostic of their character traits. This is consistent with what attribution theorists have noted: Some behaviors are more diagnostic of an individual's character than others. Reeder and Brewer (1979), for instance, argued that some dimensions of behavior are asymmetrically informative about the character of the actor. This asymmetric diagnosticity is especially true for moral behaviors. For example, dishonest people often tell the truth, but genuinely honest people rarely lie (Reeder & Brewer, 1979; Reeder, Pryor, & Wojciszke, 1992; Schneider, 1991). Likewise, violating a strong moral obligation such as not to cause harm is perceived as more reflective of personal dispositions than violating a weaker moral obligation (e.g., failing to help; Trafimow & Trafimow, 1999). This is consistent with research demonstrating the side-effect effect: Obligations to prevent foreseeable harms are treated differently than obligations to help; bringing about harm that one foresees is therefore perceived as intentional; and such acts are seen as more informative about a person's character.

More generally, from a character-based perspective it makes sense to hold someone fully accountable—that is, to treat his or her actions as though they were intentional—for a decision made when they could foresee that it would cause harm. Such an act sends a clear signal as to what the agent does and does not value. And, if as we have been arguing, judgments of blame are in the service of character evaluations, there is a world of difference between someone knowingly allowing or causing harm to occur and someone taking potential harm seriously and ensuring that it does not occur.

An actor's intentions provide us with information about both the nature of act itself and the agent who performed it. Accordingly, intentions may play an independent causal role in our perception of the outcomes (Gray & Wegner, this volume). This was demonstrated recently by Gray and Wegner (2009), who found that when participants received shocks that they thought were intentional, they found them more painful than unintended shocks of equal magnitude. Moreover, whereas continued administration of unintentional shocks led to a reported decrease in the severity of pain (consistent with psychophysical laws of habituation), intentional continued to have the same perceived effects throughout the testing procedure, suggesting that the pain of intentional harm is more difficult to accommodate.

In sum, intentions matter to moral judgment above and beyond the specifics of a given action; they are important because we see ourselves and others as rational and purposeful agents, and intentions are the clearest ways of understanding what causes a person to do what he or she does; they reflect the person's attitudes, traits, and general moral character (Morse, 2003).

Beyond intentional behaviors, other kinds of behavior are also considered to be diagnostic of bad moral character—for example, acts that seem to indicate emotional callousness or a failure to consider the welfare of others when making a decision. Even when an act produces a net benefit for others, in some cases we find it difficult *not* to blame the agent. This again poses a real puzzle for standard accounts of blame, because it means that even though a person may perform a morally permissible (or even obligatory) act, and one that fails to harm others or even ends up helping others, the person may be considered blameworthy.

One way to understand this from a character-based perspective is that we do not just want individuals to perform the right act; we want them to do it in the right way and for the right

reasons. Consider the example of research on the footbridge dilemma, Thomson's (1976) scenario in which a person is faced with the decision of throwing a large man off a footbridge to his death in order to save the lives of five other people. Although most people view this act as morally forbidden (Mikhail, 2007), upon reflection many people agree that it might actually be the most ethical choice (Greene et al., 2008). But there are different ways in which the decision process can be described. In some cases, the person making the decision is described as painfully deliberating until the very last moment (with a train rapidly approaching), when he finally decides it is the right thing to do. In other cases the person making the decision immediately shoves the large man to his death while laughing. Although both men performed the same kind of act, with the same consequences — killing one to save five—it is difficult not to think that the "laughing utilitarian" deserves a negative moral evaluation. Indeed, recent research indicates that people often evaluate a person based not on the specific consequences of the act, but rather — independent of consequences — on what the act reveals about the person's character.

One piece of information often thought to be diagnostic of an agent's mental state is his or her emotional state at the time of the action. Was the person in a calm, rational state of mind, or was the person acting impulsively? Emotionally impulsive acts are generally seen as less controllable (which is one reason premeditated murders are punished more harshly than "crimes of passion"). Yet actions that are viewed as equally impulsive—where controllability is held constant—can lead to differential judgments of responsibility depending on the valence of the act. For instance, Pizarro, Uhlmann, and Salovey (2003) found that, consistent with an act-based approach to moral decision making, participants tended to reduce blame if a negative act was committed impulsively rather than deliberately. A person who impulsively hit someone in a fit of

anger was seen as less responsible than someone who deliberately decided to hit someone. However, contrary to the invariance predicted by the act-based approach, positive acts that were committed impulsively received no such reductions in responsibility compared to positive acts committed deliberately. For example, impulsively donating money to charity because of a strong sympathetic reaction did not result in lower responsibility judgments or praise than donating the same amount after having deliberated about it. The authors argued that participants were making inferences about the actors "metadesires" (the extent to which the donor had a second-order desire to entertain positive or negative impulses), and that the observed asymmetry arose because, unlike positive impulses, negative impulses were assumed to be unwanted by the participant. Consistent with this interpretation, follow-up studies revealed that when positive impulses were described as unwanted, the asymmetry disappeared. These metadesires — the evaluations an individual makes regarding his or her first-order impulses — are indicators of what the person truly values, or of the "deep" self (Wolf, 1996).

Woolfolk, Doris, and Darley (2006) found that actors can sometimes be judged as morally responsible even if their actions were completely constrained by external circumstances. According to act-based models, acts committed because of situational constraints (indicating less controllability over the act) should cause research participants to reduce perceived responsibility for the action. But when Woolfolk and colleagues presented a scenario in which a man was under a clear situational constraint that forced him to murder an airplane passenger (he was forced by hijackers to kill the person or else he and 10 others would be killed), they held him responsible for the murder if it was something he had wanted to do anyway (that is, if he "identified" with the act). On the other hand, if participants believed that, while under identical situational

constraints, the agent did not identify with the action—that in some sense the behavior felt "alien" to him—they reduced their attributions of responsibility. On the standard account of moral reasoning this is an anomaly, but on the character-based account it is quite reasonable. Embracing, or identifying with, murderous behavior is diagnostic of an individual's character.

But not all impulsive acts provide the same information about an individual's character. Critcher, Inbar, and Pizarro (under review) found that only certain kinds of negative impulses led to a reduction in blame compared to identical deliberate acts. Some negative impulses actually led to increased blame. Consistent with the findings of Pizarro et al. (2003), a negative behavior committed while an actor was enraged resulted in lower blame than a similar but deliberate act. However, negative acts committed in a "rash" manner—equally impulsive, but without the presence of strong negative emotions—led to amplified blame when compared to a deliberate action. Critcher et al. found that this effect resulted from differing assumptions regarding the metadesires of individuals who act impulsively due to "rashness" compared to those who act impulsively out of "rage." Consistent with the character-based approach to moral reasoning, this effect was shown to occur because acts of rashness are perceived as more diagnostic of the underlying intentions and character attributes of the individual than acts of "rage," and are perceived as less situationally determined than acts of rage.

In another set of studies that resemble the "laughing utilitarian" example discussed earlier (Critcher, Helzer, Tannenbaum, & Pizarro, in preparation), we demonstrated another way in which acts that result in the same — or in some cases better — consequences can be seen as blameworthy. We showed that the manner in which the decision is carried out—not just the decision itself—affects judgments of praise and blame. To the extent that the materials presented

to participants provided information about the agent—what he or she valued and knew about what other people value—research participants incorporated these cues into their moral judgments. In one study, we presented participants with a common moral dilemma in which a group of Jewish people must remain quietly hidden lest nearby Nazi soldiers hear them and kill them. In this dilemma, a crying baby must be silenced or the group will be discovered and killed. The group realizes that the only way to save everyone's life is to kill the crying baby (by suffocating him). One group of subjects receives the information that the Nazi soldiers are next door, and a decision about the baby must be made quickly, while the other group is told that the Nazis are a few houses away and the potential victims have time to deliberate. When told that the person in charge of making the decision chose to sacrifice the baby, participants judged him significantly more harshly if he made this decision immediately than if he made it after having had a chance to deliberate. In fact, choosing to sacrifice the baby immediately garnered the most negative moral evaluations, and the other three conditions were evaluated less negatively and to the same extent. When making a difficult decision about morality in a situation like this, it appears that people want the decision to be made with difficulty, because this indicates that the decision maker has sentiments we value.

Finally, consider another set of studies by Tannenbaum and colleagues (Tannenbaum, Uhlmann, & Deirmeier, under review) showing that evaluations of character can result in greater blame for acts that cause *less* harm. In one of their experiments, participants were given one of two descriptions of a company manager who causes harm to his employees (by cutting their vacation days in half). In one condition described a "misanthropic" manager who cuts vacation days for all of his employees. In another condition he was described as only cutting vacation

days for his African American employees ("bigot" manager). (In both cases, the description stated that about 20% of the company's employees were African American, and in the bigot version only the manager knew that some employees received less vacation days). Not surprisingly, although the bigoted manager caused material harm to far fewer individuals, he was judged as more blameworthy than the misanthropic manager. Moreover, participants were more likely to believe that the behavior of the bigot was diagnostic of his character than the behavior of the misanthrope. Of importance for the character-based theoretical perspective, judgments of the diagnosticity of the bigot's behavior were significantly correlated with judgments of blameworthiness. Once again, judgments of character seemed to affect moral blame in a manner inconsistent with act-based approaches to moral reasonsing but quite consistent with a character-based approach.

Conclusion

A growing body of evidence suggests that the ways in which people make attributions of control, intentionality, responsibility, and blame are more complex and potentially important than one would assume based on an act-based model of moral judgment. Although traditional act-based approaches recognize that these criteria are sometimes important for determining blame, they fail to explain *why* they are important. We have argued that their importance is based on their being informative in that they indicate who the actor is and what he or she values and considers when performing morally relevant actions. In short, they reveal an agent's moral character, and data from several studies indicate that character is an important consideration when people assess others on moral grounds. Moreover, given the ability of the character-based

approach to moral judgments to explain findings that seem puzzling from an act-based perspective, there is reason to take this approach seriously.

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.

Ames, D. R., & Johar, G. V. (2009). I'll know what you're like when I see how you feel. *Psychological Science, 20,* 586-593.

Anscombe, G. E. M. (1958). Modern Moral Philosophy, *Philosophy, 33*, 1-19.

Aristotle (4th Century, B.C.E./1998). *The Nicomachean ethics.* Oxford: Oxford University Press.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6,* 269-278.

Bayles, M. (1982) Character, purpose and criminal responsibility. *Law and Philosophy, 1,* 5-20.

Chang, L. J., & Sanfey, A.G. (2009). Unforgettable Ultimatums? Expectation violations promote enhanced social memory following economic exchange. *Frontiers in Behavioral Neuroscience, 3,* 1-12.

Critcher, C., Helzer, E., Tannenbaum, D., & Pizarro, D. A. (in preparation) Moral judgments stem not from the goodness of acts, but from the goodness of the principles motivating them.

Critcher, C., Inbar, Y., & Pizarro, D. A. (under review). When impulsivity illuminates moral character: The case of rashness.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). First judge warmth, then competence: Fundamental social dimensions. *Trends in Cognitive Sciences, 11,* 77-83.

Frank, R. H. (1988). *Passions within Reason: The Strategic Role of the Emotions.* New York: Norton.

Foster, E. K. (2004). Research on gossip: Taxonomy, methods, and future directions. *Review of General Psychology, 8,* 78-99.

Gintis, H., Henrich, J., Bowles, S., Boyd, R., & Fehr, E. (2008). Strong reciprocity and the roots of human morality. *Social Justice Research, 21*, 241-253.

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science, 19,* 1260-1262.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107,* 1144-1154.

Greene, J. D., Nystrom, L. E., Engell, A. D., Darley, J. M., & Cohen, J. D. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron, 44,* 389-400.

Greene, J. D. (2007). Why are VMPFC patients more utilitarian?: A dual-process theory of moral judgment explains. *Trends in Cognitive Sciences, 11,* 322-323.

Grueneich, R. (1982). The development of children's integration rules for making moral judgments. *Child Development, 53,* 887-894.

Haidt, J . (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108,* 814-834.

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

Hart, H. L. A. (1968) *Punishment and social responsibility.* Oxford, UK: Clarendon Press.

Heider, F. (1958). *The psychology of interpersonal relations.* New York: Wiley.

Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009) Disgust sensitivity predicts intuitive disapproval of gays. *Emotion, 9,* 435-439.

Kagan, S. (1998). *Normative ethics*. Boulder, CO: Westview Press.

Kant, I. (1796/2002). *Groundwork for the metaphysics of morals* (tr. Arnulf Zweig). New York: Oxford University Press.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation* (pp. 129-238). Lincoln: University of Nebraska Press.

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

Knobe, J., & Doris, J. (in press) Strawsonian variations: Folk morality and the search for a unified theory. In J. Doris (Ed.) *The Handbook of Moral Psychology*. Oxford, UK: Oxford University Press.

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.

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

Mikhail, J. (2007). Universal moral grammar: Theory, evidence, and the future. *Trends in Cognitive Sciences, 11,* 143-152.

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

Monin, B., Pizarro, D., & Beer, J. (2007). Deciding vs. reacting: Conceptions of moral judgment and the reason-affect debate. *Review of General Psychology, 11,* 99-111.

Morse, S. J. (2003). Diminished rationality, diminished responsibility. *Ohio State Journal of Criminal Law, 1,* 289-308.

Pizarro, D. A., & Helzer, E. (in press) Freedom of the will and stubborn moralism. In R.F. Baumeister, A. R. Mele, and K. D. Vohs (Eds.), *Free will and consciousness: How might they work?* New York: Oxford University Press.

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

Pizarro, D. A., Uhlmann, E., & Salovey, P. (2003). Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science, 14,* 267-272.

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.

Rand D. G., Dreber A., Ellingsen, T., Fudenberg, D., & Nowak M.A. (2009). Positive interactions promote public cooperation. *Science, 325,* 1272-1275.

Reeder, G. D,, & Brewer, M. (1979). A schematic model of dispositional attribution in person perception. *Psychological Review, 86,* 61-79.

Reeder, G. D., Pryor, J. B., & Wojciszke, B. (1992). Trait-behavior relations in social information processing. In G. R. Semin & K. Fielder (Eds.), *Language, interaction, and social cognition* (pp. 37-57). Newbury Park, CA: Sage.

Schneider, D. J. (1991). Social cognition. *Annual Review of Psychology, 42,* 527-561.

Searle, J. (1983). *Intentionality*. Cambridge, UK: Cambridge University Press.

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

Smart, J. J. C., & Williams, B. (1973). *Utilitarianism: For and against.* Cambridge, UK: Cambridge University Press.

Solum, L.B. (2004) The Aretaic Turn in Constitutional Theory. University of San Diego Legal Working Paper Series. University of San Diego Public Law and Legal Theory Research Paper Series. Working Paper 3.

Tannenbaum, D. T., Uhlmann, E. L., & Diermeier, D. (under review) Moral signals, public outrage, and immaterial harms.

Thomson, J. (1976). Killing, letting die, and the trolley problem. *The Monist, 59,* 204-217.

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

Trafimow, D., & Trafimow, S. (1999). Mapping imperfect and perfect duties on to hierarchically and partially restrictive trait dimensions. *Personality and Social Psychology Bulletin, 25,* 686-695.

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

Wellman, H. M., & Miller, J. G. (2008). Including deontic reasoning as fundamental to theory of mind. *Human Development, 51,* 105-135.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science, 17,* 592-598.

Wolf, S. (1987). Sanity and the metaphysics of responsibility. In F. Schoeman (Ed.), *Responsibility, character, and the emotions: New essays in moral psychology* (pp. 363-373). Cambridge, UK: Cambridge University Press.

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