Metaskepticism About Moral Responsibility

Abstract:
Contemporary philosophical theories of moral responsibility attempt to develop universal criteria for fair assignments of blame and praise. The theories also appeal to intuition about key principles and cases to justify these criteria. Consequently, theories of moral responsibility must make empirical assumptions about the universality or convergence of the intuitions upon which their theories rely. This paper lays out a comprehensive empirical and philosophical challenge to these assumptions. I present evidence suggesting that there are *essential and incompatible* intuitive differences regarding the conditions for fair assignments of moral responsibility. I then argue that these differences are sufficiently deep and well-motivated to make it implausible that reflection, concept disambiguation, dialogue, and agreement about non-moral facts could resolve them. I conclude therefore that we have no principled means of establishing the truth of any contemporary theory of moral responsibility. I call the resulting position 'metaskepticism about moral responsibility' because the challenge applies not only to positive accounts of moral responsibility, but also to skeptical theories which claim that human beings *everywhere* cannot deserve praise or blame for their actions.

1. Introduction

Just after the tragedy at Virginia Tech University, a reporter from the NPR program "Day to Day" set out to interview Koreans living in Los Angeles about the massacre. At first the reporter had trouble finding anyone who was willing to answer her questions. Some actually fled from the microphone. Finally, a Korean realtor agreed to be interviewed. He claimed to be deeply ashamed about the incident. The reporter was incredulous: "Why? You had nothing to do with it!" The man replied: "I know, but he was a fellow Korean."[1]

In the same week Rev. Dong Sun Lim, founder of the Oriental Mission Church in Koreatown, released this statement: "All Koreans in South Korea – as well as here – must bow their heads and apologize to the people of America." And South Korean Ambassador Lee Taesik called on Korean Americans not just to be ashamed, but to repent. He suggested a 32-day fast: one day for each victim of the carnage.

[1] "U.S. Koreans Respond to Shootings" *Day to Day*, April 18, 2007.

Many Americans found this attitude baffling. Why should Koreans living thousands of miles away from Blacksburg, Virginia feel compelled to apologize (never mind starve themselves) for something over which they had no control? What do they have to apologize *for*? Adrian Hong, a board member of the Mirae Foundation, a national organization of Korean-American college students, offers this explanation: "First-generation Koreans tend to have a cultural sense of shared responsibility. If something good happens to one, it happens to all Koreans, and if something bad happens to one, it happens to all of them." UCLA Anthropology Professor Kyeyoung Park adds: "In Western culture there is an emphasis on guilt; in many Eastern cultures the emphasis is on shame. I think Korean-Americans want to do something [about the incident] because they feel ashamed. Some of them feel *truly responsible*, even though it is ridiculous to think they are responsible for the action of this person." [my italics][2]

The Koreans' sense of shared blame, along with the failure of many Americans to understand this feeling, is just one example of the fundamentally different perspectives about moral responsibility that may be found across cultures. At bottom, these differences concern beliefs about the conditions or criteria for fair assignments of blame and praise. The incredulity of the 'Day to Day' reporter ('but you had nothing to do with it!) illustrates a common Western intuition that in order to be genuinely blameworthy for a state of affairs, you must have played a role in bringing it about. This intuition is so deeply embedded in the Western belief system that it appears self-evident, like a mathematical truth or an elementary rule of logic. This paper will argue, however, that this intuition, and others relating to moral responsibility, are not as universal as they appear.

---

[2] Citations from "Korean-American groups express sorrow, avoid guilt: Korean-Americans burdened with guilt, shame and fear of backlash," by Kelly Brewington. *Baltimore Sun*. April 21, 2007.

These differences are not merely interesting from an anthropological perspective. They are also philosophically significant—deeply relevant, I will argue, to the contemporary debate about moral responsibility. This is because (1) contemporary philosophical theories of moral responsibility develop universal conditions for fair assignments of blame and praise, and (2) they appeal to intuition to justify these conditions. Consequently, the theories must make empirical assumptions about the universality or convergence of the intuitions to which they appeal. This paper will lay out a comprehensive empirical and philosophical challenge to these assumptions. I present evidence suggesting that there are *essential and incompatible* intuitive differences regarding the conditions for fair assignments of moral responsibility. I then argue that these differences are sufficiently deep and well-motivated to make it implausible that reflection, concept disambiguation, dialogue, and agreement about non-moral facts could resolve them. I conclude therefore that we have no principled means of establishing the truth of any contemporary theory of moral responsibility.

This challenge applies not only to positive accounts of moral responsibility, but also to skeptical theories, views that human beings *everywhere* cannot deserve praise or blame for their actions. Consequently, I call the position defended in this paper *metaskepticism about moral responsibility*.[3]

2. The Appeal to Intuition

I should be clear from the outset that moral responsibility is here referred to in the strong "desert-entailing sense" (Strawson, 1986). To believe that someone is morally responsible for an action in the sense relevant to this paper is to believe that the person *deserves* blame or praise—in other words that the person ought to be blamed or praised (and perhaps punished or

[3] This paper owes a great deal in the structure of its argument to Nichols, Stich, and Weinberg (2003).

rewarded) independent of any consequentialist benefits that might arise from these assignments of responsibility. Although the term 'moral responsibility' is used in many other ways, it is the non-consequentialist, desert-entailing sense of responsibility that is at the center of the philosophical debate.[4]

Arguments for incompatibilism about moral responsibility—the view that the truth of determinism would make moral responsibility impossible—generally employ at least one of two key incompatibilist principles: (1) the Principle of Alternate Possibilities (PAP) which states that we cannot be morally responsible for an act if we lack the ability to do otherwise; and (2) the 'Transfer of Non-Responsibility' (TNR) principle, which states (roughly) that we cannot be morally responsible for an act if we are not morally responsible for any of the determining factors of that act.[5] (The non-responsibility for the determining factors 'transfers' to the act itself.) Incompatibilists either argue directly for the intuitively plausibility of these principles (e.g. van Inwagen, 1983), or they describe specific cases in which an agent is intuitively not morally responsible and then argue that there is no relevant difference between those cases and *all* instances of fully determined human behavior.[6] These "generalization strategies" (as R.J. Wallace has called them) cannot get off the ground, however, unless the reader shares with the author the starting intuitions that the agents in the original cases are not morally responsible for

---

[4] There is near universal agreement about whether we can be morally responsible in other ways. Libertarians, compatibilists, and skeptics alike believe that blame, praise, punishment, and reward can have consequentialist benefits. Nor is there much disagreement over whether we have compatibilist forms of freedom and responsibility The debate centers on the question of whether compatibilist freedom is sufficient to ground desert-entailing moral responsibility.

[5] Fischer and Ravizza (1998) provide the following more careful formulation of this principle: "(1) p obtains and no one is even partly morally responsible for p; and (2) if p obtains, then q obtains, and no one is even partly morally responsible for the fact that if p obtains, then q obtains; then (3) q obtains, and no one is even partly morally responsible for q." (p. 152) .

[6] Pereboom's (2001) 'four case argument,' for example, defends the incompatibilist intuition that one can only be morally responsible for an action if it is not produced by a deterministic process that traces back to factors beyond the agent's control (a rough approximation of the TNR principle). Pereboom does this by describing cases in which the reader judges that an agent is not morally responsible and shows how those cases are, in the sense relevant to moral responsibility, identical to all cases involving determined human action.

their behavior. Readers who do not share those intuitions will not arrive at the incompatibilist conclusion.[7]

Intuitive plausibility is no less relied upon in the cascade of counterexamples which compatibilists have developed to undermine incompatibilist principles. Frankfurt (1969) provides the most famous of these in his attack on the Principle of Alternate Possibilities (PAP). Frankfurt argues that accepting PAP entails a counterintuitive conclusion—namely, that an assassin who carried out his intention of killing a prime minister is not morally responsible for that act (because he had a chip in his head that guaranteed that he would commit the act even if he did not have the intention). My aim is certainly not to take sides on the question of whether or not this is an effective counterexample; it is only to point out that *if* it is an effective counterexample, it must be so in virtue of how well it accords with the reader's intuitions about the blameworthiness of the assassin. Frankfurt-style examples do not reveal a logical contradiction in the incompatibilist's position, nor do they reveal any false empirical assumptions. It is *logically* open for the incompatibilist to respond: "well, the assassin could not have done otherwise, therefore PAP is violated and he is not morally responsible—I stand by that." We would deem this response not false exactly, but *disingenuous*, because even incompatibilists cannot *really believe* that under these exact circumstances the assassin is not morally responsible for his actions.[8]

---

[7] See Sommers (forthcoming) for a more detailed discussion of how incompatibilists must appeal to intuition to support their arguments.

[8] Fischer and Ravizza employ the same strategy in their arguments against the TNR principle (see Ravizza (1994), Fischer and Ravizza (1998); and the "erosion" and erosion* cases). The structure of these arguments are as follows:

1. The TNR principle entails that Agent P would not be morally responsible for act A, in circumstance C.
2. But it is intuitively obvious that that P *is* morally responsible for A, in C.
3. Therefore the TNR principle is false or implausible.

Again, if we believe these counterexamples to be effective, or even worth responding to, it is only because we share the intuition that the agent is indeed morally responsible in the circumstances described by Fischer and Ravizza.

Compatibilist counterexamples may also lead to further refinements of compatibilist theories of moral responsibility. But (yet again) we cannot evaluate whether the refined theory is truly improved without consulting our intuitions about moral responsibility in key cases. A representative example is Susan Wolf's criticism of what she calls the 'deep self view"—the compatibilist position that one is morally responsible for an act that reflects one's authentic self or character.[9] Wolf begins her criticism by offering a counterexample. JoJo is the son of an evil sadistic dictator, Jo the First. Jo the First has trained JoJo from early childhood to value arbitrary expressions of cruelty, such as executing or torturing his subjects on the basis of mere whim. JoJo understandably sees his father as a role model and acquires a fully authentic deep self that values and endorses cruel behavior as well. According to the 'deep self view,' JoJo is just as blameworthy as someone who had the kindest and most conventionally moral upbringing. But according to Wolf, "in light of JoJo's heritage and upbringing, it is dubious at best that he should be regarded as responsible for what he does." (Wolf, 2003, pp. 379-380). So Wolf adds 'the condition of sanity' to the deep self view, arguing that if actions are a reflection of the agents' authentic selves, *and* the agents are able to understand the difference between right and wrong (which JoJo lacks), then is the agent is morally responsible.

Wolf's theory depends on two premises that require intuitive agreement. First, we need to share the (perhaps controversial) intuition that JoJo is not morally responsible for his cruel behavior. Second, we must believe that the addition of the 'sanity requirement' provides the deep self view with sufficient conditions for moral responsibility. (And of course, to argue

[9] Frankfurt (1971) and Watson (1975) are Wolf's primary targets.

against Wolf's theory one must devise an intuitively plausible counterexample. The cycle continues.[10])

In sum, all existing theories of moral responsibility appeal to intuitions about key premises and cases. True, we may be moved to adjust or reconsider or revise these intuitions if we are persuaded by certain arguments or learn more about non-moral facts, such as the way human decision-making actually functions. But ultimately our intuitions—or "considered judgments," or however one wishes to describe reflective core beliefs about cases or principles—play an essential role in our acceptance or rejection of the arguments for each particular theory.

3. The Empirical Challenge to "Universalist" Theories of Moral Responsibility

Contemporary theories of moral responsibility have another feature in common: the conditions or criteria for moral responsibility contained in the theories are designed to apply universally, for all agents, for all societies. By this I don't mean that the theories conclude that all (or even any) adults in a given society can be morally responsible for their behavior. It may be that no member of a particular culture can meet the theory's criteria for justified assignments of blame and praise. But the conditions or criteria *themselves* are meant to apply across cultures in such a way that *if* the sufficient conditions specified in the theory are met, the agents are judged to be morally responsible; and if a necessary condition is not met, the agent judged to be exempt from morally responsibility (although not necessarily exempt from punishment). I refer to theories that have these universal criteria as 'universalist' theories of moral responsibility.[11]

---

[10] See Knobe and Doris (forthcoming) for a discussion of how the cycle of counterexamples from both sides lead to what John Fischer has called a "dialectical stalemate."

[11] I should distinguish universalist theories from what Knobe and Doris (forthcoming) call 'invariantist' theories of moral responsibility. Knobe and Doris argue theories of moral responsibility are invariantist, in that they include the common assumption that "people should apply the *same* criteria in *all* of their moral responsibility judgments." They believe this assumption is mistaken. Their focus is on *intrapersonal* variation of intuitions about moral responsibility; they believe that an individual's criteria vary according to the particular context of the judgment.

Because they contain universal conditions for genuine blameworthiness and praiseworthiness *and* appeal to intuition for the justification of these conditions, theories of moral responsibility are vulnerable to an empirical challenge: in particular, the claim that the intuitions to which the theories appeal are not nearly as universal as the theories (at least implicitly) presume. This challenge is in many ways a version of the argument from relativity or disagreement.[12] There are some differences worth mentioning, however. Arguments from disagreement in the metaethical literature focus on objective moral values in general; my argument only addresses values relating to moral responsibility. Second, arguments for moral realism or objectivism in general may not rely as much on appeals to intuition as arguments for various positions on moral responsibility. Finally, whereas arguments from disagreement in metaethics are often presented in support of moral skepticism, my challenge applies equally well to skepticism about moral responsibility.

The empirical challenge has two parts. The first is to show that intuitions about the conditions for justified assignments of moral responsibility radically differ across cultures. The second is to show that these differences are fundamental; in other words, that they unlikely to be resolved by philosophical analysis and reflection, using a method such as 'wide reflective equilibrium' (Rawls, 1971; Daniels, 1979). I begin the first part by reviewing evidence that some cultures may not have anything like a recognizable *control condition* for moral

---

Since my focus is on interpersonal variation, universalist theories differ from 'invariantist' theories in the following sense. A theory may be invariantist but not universalist, if it is only meant to apply to a specific culture. For example, one might propose an invariantist necessary condition for judging blameworthiness in every action—say reason-responsiveness—but one that is only meant to apply within a particular culture (say, 'reason cultures.') Conversely, a theory can be variantist but universalist in that it is meant to be true for all cultures. A variantist theory might lay out certain criteria for the blameworthiness of husbands (which is different from the criteria that apply to the blameworthiness of strangers), but claim that these criteria apply in all cultures whenever the object is the husband. Indeed, I believe Knobe and Doris lean towards a universalist version of variantism. (As does Smilansky (2000)'s "dualism," another version of variantism.) So, when I say that a theory of moral responsibility is universalist, I mean only that the criteria, invariantist or variantist, are meant to apply universally, across cultures.

[12] See Mackie (1977) and Loeb (1998) for two recent examples of arguments from disagreement. And See Doris and Stich (2005) and Doris and Plakias (forthcoming) for excellent discussions of the role of empirical research in these arguments.)

responsibility. Virtually all Western theories of moral responsibility contain some type of control condition, although the theories differ on what it means for an agent to have the right kind of control over an action. Theories that lacks a control condition are referred to as 'strict liability'—individuals may be blamed or punished whether or not there was intent, *mens rea,* or negligence on the individual's part. Most people in the West regard strict liability approaches as fundamentally unfair, although they may be invoked under certain occasions for consequentialist reasons. As Thomas Nagel puts it: "strict liability may have its legal purposes, but seems irrational as a moral position." (Nagel, 1979, p.31.)

Evidence from cultural anthropology suggests, however, that denying the control condition for moral responsibility may only "seem irrational" to people in certain kinds of societies. Research on what have been come to be called honor cultures, for example, suggests that agents do not need to have any kind of control over an act in order to be deemed fair and morally appropriate targets of punishment. This is evident by the common practice of collective punishment—retaliation against people who played no part in committing the offense that is to be avenged. In honor cultures it is quite common to target friends, family members, or associates of the offender. Consider this description of the Netsilik Eskimos:

> The objective of the revenge party was not just to kill the original murderer but members of his kindred as well. In a sense the members of the kindred *shared responsibility* of the murder. (Balicki, 1970, p.182)

Individuals seem to be morally responsible in the full blooded desert-entailing sense of the word for the actions of their friends and relatives. This is a radically different way of viewing the criteria for moral responsibility than those commonly found in the West. No matter how one

interprets the control condition, it cannot involve merely being the cousin or friend of someone who performs the act.[13]

One might object that the retaliators in these kinds of cases do not truly deem their targets blameworthy or morally responsible. In the case of collective punishment, for example, perhaps the offended parties are trying to punish and hurt the real offender by harming his friends and family. The problem with this explanation is that these attacks occur even if the offender is dead (and so cannot be harmed by the death of a family member). So perhaps the retaliators are trying to send a message to deter future attacks; perhaps they regard the non-offending targets much like Americans view enemy civilians in a just war—people who are (unfortunately) caught in the crossfire. However, while deterrence and other pragmatic goals may be part of the function of these kinds of attacks, it does not seem to be the whole story. Retaliation in honor cultures is never described as 'collateral damage' or 'regrettable but necessary.' Members of honor cultures feel no need for such consequentialist justifications. There is no acknowledgment that the non-offenders are being unfairly punished. On the contrary, after an attack the offended individual is thought to have a *duty* to retaliate against suitable targets, which explicitly include individuals whose only crime is being related to the offender in a certain way. It is deemed perfectly fair to suffer punishment for the acts of relative.

We find another example of an absence of a control condition for responsibility in the terrible and almost incomprehensible practice of honor killings. (Indeed, the incomprehensibility of this practice from a Western perspective suggests that there may be tremendous variation in

[13] See Boehm (1984), Miller (1990, 1994), for many other examples of 'shared responsibility.' See also Elster (1991) for an excellent discussion of honor norms. And see Feinberg (1968) for an analysis of collective punishment.

intuitions about moral desert.[14]) The term 'honor killing' refers to the murder of a woman or girl (usually by a family member) because she has lost her virginity—a stain on the family's honor. What makes these cases even more unfathomable is that the murders occur even when the woman is a victim of rape. In the West, we find this perspective not just brutal and cruel but bewildering. Even if one believes that premarital sex is a mortal crime, how could it possibly be justifiable to punish rape victims? The woman clearly does not have any kind of control—compatibilist or incompatibilist—over the loss of her virginity. Yet the woman is still judged to be culpable, and therefore an appropriate target of punishment in the eyes of her family and even the legal system. According to the Middle East Report (Spring 1998), "maintaining honor is deemed a woman's responsibility, *whether or not she has been educated about sex or consented to the act."* (my italics).[15] It is true that in some cases the family members view wives and daughters exclusively as property.[16] But in other cases, the victim is a *beloved* daughter or sister. The murderer may be consumed with grief over the act, yet he seldom repents. The phenomenon becomes a little easier to understand (although not to condone) if we imagine that the cultures that support this behavior do not see control as being a necessary condition for moral responsibility. If that is the case, it becomes almost irrelevant that the victim did not choose, or intend, to lose her virginity. What matters is that she did lose her virginity, bringing dishonor on the family. This is why tremendous grief is consistent with a complete lack of repentance about

[14] See Bowman (2006) for a similar sentiment. Bowman argues that the "sheer strangeness [of honor killings] ought to alert us to the difficulty of understanding cultural honor where it exists." (pp. 17-18)
[15] Consider too this example of institutionalized punishment of rape victims: "In Saudi Arabia, a 19-year-old girl who was kidnapped at knifepoint, gang-raped, and then beaten by her brother for having "allowed" herself to become the victim of a rape has been sentenced to 90 lashes. Her crime? Indeed, one of her judges told this young woman she was lucky to have not gotten jail time."
[16] Thanks to Theresa Lopez for raising this complication about the attitudes of individuals in honor cultures towards woman.

the killing. What we consider to be the terrible *injustice* of these killings would not resonate to members of these cultures nearly as much as it does to those of us in non-honor cultures.[17]

It is possible that those who perform this act do so not out of any sense that the woman deserves it, but purely out a higher duty to rescue the honor of the family. Perhaps the rape is viewed as some kind of terminal disease.[18] On the other hand, there are cases in which the punishment of the rape victim is not death but lashes; from ruling issued by judges! These rulings indicate that there is some notion of moral desert at work, although it may be argued that the blame stems from some exotic interpretation of the woman's intent. Certainly, there is more to discuss about the possible explanations for these practices. At the very least, however, evidence about honor killings as well as collective punishment show that we must ask these questions in order to understand the intuitions and moral perspectives of those who engage in these practices.

One finds additional evidence of deep intuitive differences about moral responsibility from research on what are termed 'shame cultures' (in contrast with 'guilt cultures') and "collectivist societies" (in contrast with 'individualist societies'). Members of shame cultures feel more responsible for conspicuous acts that are clearly not of their own doing, as illustrated in the example of Koreans feeling morally responsible for the act of the Virginia Tech killer. In guilt cultures, individuals are more likely to find an act like this appalling but not something they personally deserve blame for in any way. In addition, members of shame cultures appear to feel less responsible for a class of actions that would count as morally responsible behavior in guilt cultures. Benedict (1947), for example, argues that moral norms in shame cultures depend in

---

[17] See also Williams (1993) and his discussion of the Greek notion of responsibility, which also extends "beyond our normal purposes and what we intentionally do." (p. 74).

[18] According to one tribal leader, "a woman is like an olive tree. When its branch catches a woodworm, it has to be chopped off so that society stays clean and pure." (Clark 2003)

large part on the *perception* of one's behavior, character traits, and appearance. "True shame cultures," she writes, "rely on external sanctions for good behavior, not, as true guilt cultures do, on an internalized conviction of sin." (Benedict, 1947, p. 223.) Guilt cultures, by contrast, focus on the act from the individual's perspective:

> [Shame] requires an audience or at least the fantasy of an audience. Guilt does not. In an nation where honor means living up to one's own picture of oneself, a man may suffer from guilt though no man knows of his misdeed and a man's feeling of guilt may actually be relieved by confessing his sin. (Benedict, 1947, p. 223)

In his book *Culture's Consequences* Hofstede argues that the shame/guilt culture distinction applies to collectivist and individualistic societies as well. Hofstede cites E.M. Forster's *A Passage to India* to contrast the motivations of people from these two types of cultures:

> Aziz upheld the proprieties, though he did not invest them with any moral halo, and it was here that he chiefly differed from the Englishman. His conventions were social. There is no harm in deceiving society as long as she does not find you out, *because it is only when she finds you out that you have harmed her*. (Hofstede, 1980/2001, p. 212, my italics).

In somewhat oversimplified terms, if agents violate norms in a shame culture but the violation is not discovered, the agents are less likely to hold themselves responsible. Agents in guilt cultures will hold themselves responsible whether or not the offense is discovered. One famous example of 'guilt culture responsibility' is Raskolnikov in *Crime and Punishment*. He feels searing guilt and a subsequent need to confess to a crime that no one knows he committed. Someone in a true shame culture, according to Benedict's and Hofstede's accounts, may find Raskolnikov's inner turmoil more difficult to understand.

Less striking but just as suggestive are data that come from the recent studies of Japanese (collectivist) and American (individualist) intuitions on moral responsibility. Hamilton and Sanders (1981, 1983) presented scenarios to subjects from Detroit and two sets of subjects in Japan, and asked them to assess the moral responsibility of the offenders. The scenarios were

manipulated according to the mental state of the defender, and the authority and status of the offender and victim. The authors report that:

> Japanese respondents made significantly less use of mental state information than Americans in both surveys, made more use of information regarding the hierarchy of roles; more regarding role solidarity; and more regarding influence from another. (Hamilton and Sanders, 1983, p. 208).

According to the authors, it matters less to subjects in collectivist societies whether the offender intended to perform the act. Furthermore, to the extent that it does matter, the information about intentionality is used less as a means of dishing out 'just-deserts,' and more to reinforce solidarity within the group, (Japanese subjects are also far less likely to use retribution as a justification of punishment (Hamilton and Sanders, 1988).)

We have reason to think, then, that individuals in shame/collectivist cultures and guilt/individualist cultures have very different perspectives on issues relating to moral desert. Of course, these differences are a matter of degree and not kind. But it is certainly plausible that deep variation on the emphasis of shame in one culture, guilt in another, as well heavy emphasis of group solidarity in one culture, individual achievement in the other, would result in fundamentally different intuitions, norms, and beliefs about moral desert.

Of course, one might concede that there are deep intuitive differences about moral responsibility but claim that the differences can eventually be resolved by employing a method like wide reflective equilibrium. According to this view, the existing intuitive disagreement is likely a result of irrationality, superstition, and ignorance about non-moral facts. "Fully informed" individuals, stripped of irrational biases, employing wide reflective equilibrium, would converge on the same judgments and attitudes regarding moral responsibility. After all, it is likely that many honor cultures have different intuitions and beliefs about geology and biology, but this does not mean that there is no correct answer to questions in those fields. Just

as cultures have irrational views about the Earth's origin, many cultures may have irrational or false views about moral responsibility and desert.

The appeal to wide reflective equilibrium is, I believe, the most promising strategy available for defending universalist theories of moral responsibility. Note, however, that those who employ this defense are making empirical assumptions as well, assumptions about the uniformity of human psychology. Specifically, the assumption is that once all non-moral facts are agreed upon, and unfounded superstitious notions are eliminated, human beings will come to share ethical intuitions about the criteria for justified assignments of moral responsibility—whatever their physical and social environment. The second part of the empirical challenge, then, is to raise doubts about the truth of this assumption.

There are several ways to develop this part of the challenge. The first is to identify cultures with well-developed but quite different beliefs about moral responsibility and show that the differences do not essentiality involve irrationality, ignorance, or concept ambiguity on either side. From the perspective of a contemporary Westerner the responsibility norms in Saga Iceland, for example, seem arbitrary and often unjust in the extreme. But it is not at all clear how one could demonstrate that the norms are irrational. The norms and beliefs of Icelandic honor cultures are elaborate and complex; they are the subject of endless discussion, analysis, and revision from within.[19] The norms do not seem to result from superstition or incomplete knowledge about the relevant non-moral facts. What grounds do we have for thinking that more rational reflection on one side or the other would result in common intuitions or "considered judgments" about the conditions for responsibility?[20]

[19] See Miller (1990, 1993) for descriptions of these norms.
[20] One might reply that we have good grounds for this belief since the norms of those particular cultures no longer exist. By most accounts, however, the demise of the cultural perspectives of Saga Iceland was due to the influence

Another way to develop this aspect of the challenge is to examine the *origins* of intuitive differences about moral responsibility. Richerson and Boyd (2005), for example, introduce their gene-culture co-evolution model with a description of the radically different norms found in honor cultures. They discuss Nisbett and Cohen's (1996) hypothesis that a strong disposition towards protecting one's honor and reputation, and avenging even minor insults, might be well suited for relatively lawless environments—where raids are common and a single occasion of theft is capable of destroying an individual's entire wealth. Similarly, it is likely that certain features of the environment make an emphasis on individual moral responsibility more adaptive.[21] Boyd and Richerson's model can account for norm-governed behavioral and even *physiological* variation across cultures over relatively short time spans. The key point is that different environments may dictate different moral norms—norms underwritten by core physiological differences that affect emotional responses and intuitions about deservingness. If the variation of intuitions is so deep as to be grounded in our neurophysiology, which in turn is influenced by our social and physical environment, then it is unlikely that intuitive convergence concerning the criteria of moral responsibility can be reached through wide reflective equilibrium. There will always be variation in human environments.

Of course, this evidence is not decisive. It is possible that the differences in considered judgments might ultimately be resolved. It is possible that "fully informed" people, stripped of every irrational bias and religious conviction, would converge on the same judgments of moral responsibility no matter what their social and physical environment. But possible and plausible are two different things. Existing variation appears to be a natural result of the different environments in which these intuitions and judgments emerge. If norms about responsibility that

---

of Christian theology. Arguably, the intuitive convergence that has occurred is more the product of shared superstitions than rational agreement.

[21] See Sommers (2006) for a defense of this view.

lack a control condition are more suitable for a certain type of environment, then it is unclear how can we accuse individuals of irrationality *in those environments.* It may be that the intricately laid out conditions for true blameworthiness that arise from intuitions formed in Western environments simply have no place in the environments of honor cultures.

Since the claim that philosophical analysis or wide reflective equilibrium cannot resolve cross-cultural disagreement about moral responsibility is an empirical one, it may of course turn out to be false. But the evidence outlined in this paper is sufficient, I believe, to shift the burden of proof to defenders of universalist theories of moral responsibility. They must lay out and defend their empirical assumptions. As it stands, philosophers from all sides of the debate use phrases like 'intuitively we find [the agent] blameworthy' or 'it seems clear that the agent is not morally responsible,' without giving any evidence that members of different cultures would share those judgments. It is no good to say that philosophers are not in the business of providing empirical data to support their theories—that this is a job for scientists. Philosophers who defend universalist theories of moral responsibility are making scientific claims already, at least implicitly. They are making the empirical claim that intuitions about moral responsibility are, or could be, shared universally. Philosophers making claims *of any sort* are obliged to provide reason to believe that the claims are true. And the only way to judge the plausibility of empirical claims is to look at the evidence.

4. Objections/Responses

One reply to the empirical challenge described above—a common reply to arguments from disagreement—would be to claim that underneath the apparent intuitive differences, there are a set of core principles of responsibility that are universally shared. But while this response

may (or may not) have some force for other moral values, the evidence suggests that it is unlikely to succeed when the values in question concern moral responsibility. As I argue in the previous section, some cultures have a vastly stripped down control condition for responsibility—if they have one at all—which other cultures consider to be the very *essence* of moral responsibility. There is simply no existing universal agreement on core principles of moral responsibility.

Another reply would be to claim that these differences are the result of conceptual ambiguity. According to this view, anyone who believes that people can be morally responsible for an act without having brought it about in any way is guilty of a conceptual mistake; it is an *a priori* truth that moral responsibility involves control. But it is not clear how one could defend this claim without begging the question. Throughout history human beings and continue to employ the concept of moral responsibility (in the desert entailing sense) in cases where control is absent. This is evident even in certain Western cultures. A common interpretation of the doctrine of original sin, for example, is that human beings can deserve punishment for the actions of Adam and Eve, something we clearly had nothing to do with. Martin Luther explicitly defends the view that human beings can deserve damnation even though they had no control over their sinful nature. Whether or not one believes that these are *justified* assignments of desert is irrelevant to the present discussion. The point is that the concept of moral responsibility does not conceptually involve anything like a recognizable control condition.

5. Conclusion

My argument in this paper has been as follows:

(1) Universalist theories of moral responsibility cannot be justified without appealing to 'starting' intuitions about crucial premises and cases involving issues relating to moral desert.

(2) Universalist theories therefore implicitly contain empirical assumptions about the universality or convergence of core intuitions about moral responsibility.

(3) We have reason to believe that these assumptions are implausible.

(4) So we have reason to doubt the soundness of arguments for any universalist theory of moral responsibility.

One might think that my conclusion would support skepticism about moral responsibility in the same way that Mackie's argument from relativity is meant to support skepticism about objective moral values in general. The irony is that this conclusion applies equally well to *skeptical* arguments for moral responsibility, such as Galen Strawson's "Basic Argument" and Pereboom's "Four Case Argument." Like positive accounts of responsibility, skeptical theories of moral responsibility apply their criteria universally and conclude—since their criteria cannot be met—that human beings everywhere cannot deserve blame and praise. My conclusion therefore is just as skeptical of the soundness of skeptical theories as it is of positive accounts of moral responsibility.

The implication about first order theories of responsibility we *can* draw is this: we are most likely unable to establish the objective truth of any theory that contains universalist criteria for moral responsibility. So if our general concept of moral responsibility is one that features universalist criteria (another empirical question), then we have grounds to believe that concept is unjustifiable. If we have a more relativistic concept of moral responsibility, one whose criteria might vary across cultures or within cultures, then first order skepticism of our concept would

not be warranted. My untested hunch is that we do have a more universalist concept of moral responsibility, and therefore that we would have to eliminate or significantly revise our notions of moral desert in order to accommodate the metaskeptical position. (See Vargas (2004, 2005) for an example of revisionist approaches to moral responsibility.)

Finally, it is worth noting that this position has important social and political implications. A metaskeptical perspective may cure of us of the expectation that other cultures will, and should, immediately embrace Western notions of justice. An analysis of different norms about responsibility would provide a much more complete understanding of the moral motivations of members of other cultures. There would also be implications for issues of domestic social policy, criminal justice in particular. The American system of punishment is, in part, grounded in retributive principles that involve a strong notion of individual desert. The truth of metaskepticism would suggest that many contemporary theories of punishment operate according to principles that (a) have no objective foundation and (b) are not shared by individuals in many groups (honor cultures in inner cities, for example) throughout the United States. Of course, these implications require extensive analysis, and in many ways metaskepticism is best thought of as a point of departure. Where we ought to go from this point is, I believe, a fascinating question in its own right.[22]

[22] Sommers (in preparation) outlines a principled methodology of arriving at settled beliefs and practices relating to moral responsibility, given the truth of metaskepticism.

References

Balikci, A. 1970. *The Netsilik Eskimo*. The Natural History Press.

Boehm, C. 1984. *Blood Revenge.* University of Pennsylvania Press.

Benedict, R. 1947 [2005]. *The Chrysanthemum and the Sword.* Mariner Books.

Busquet, J. 1920. *le droit de vendetta et les pacii corse*. Pedone.

Daly, M. and Wilson, M. 1988. *Homicide.* Aldine de Gruyter.

Daniels, N. 1979. "Wide Reflective Equilibrium and Theory Acceptance in Ethics." *Journal of Philosophy*. 76: 256-282.

Doris, J. M., and Plakias, A. forthcoming. "How to Argue About Disagreement: Evaluative Diversity and Moral Realism." In W. Sinnot-Armstrong (ed.), *The Biology and Psychology of Morality*. New York: Oxford University Press.

Doris, J. M. and Stich, S. P. 2005. "As a Matter of Fact: Empirical Perspectives on Ethics." In F. Jackson and M. Smith (eds.), *The Oxford Handbook of Contemporary Philosophy*. Oxford: Oxford University Press.

Elster, J. 1990. "Norms of Revenge." *Ethics*, 100: 862-855.

Feinberg, J. 1968. "Collective Responsibility." *Journal of Philosophy.* Nov. pp. 674-688.

Fischer, J. and Ravizza, M. 1998. *Responsibility and Control.* Cambridge University Press.

Frankfurt, H. 1969. Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy.* 66: 829-839

Hamilton, V.L. and Sanders, J. *Everyday Justice.* Yale University Press.

Hasluck, M. 1954. *The Unwritten Law in Albania*. Cambridge University Press.

Henrich, J. et al. 2004. *Foundations of Human Sociality.* Oxford Univesity Press.

Hofstede, G. 2001. *Cultures Consequences.* Sage Publications.

Knobe, J. and Doris, J. (forthcoming) "Strawsonian Variations." to appear in J. Doris et al. *The Handbook of Moral Psychology*. Oxford: Oxford University Press.

Loeb, D. 1998. "Moral Realism and the Argument from Disagreement," *Philosophical Studies*, 90: 281-303.

Luther, M. 2007. *The Bondage of the Will.* Ambassador-Emerald International; First edition

Mackie, J.L. 1977. *Ethics: Inventing Right and Wrong.* Penguin.

Miller, W.I. 1993. *Humiliation*. Cornell University Press.

Miller, W.I. 1990. *Bloodmaking and Peacemaking.* University of Chicago Press.

Nagel, T. 1979. "Moral Luck." in *Mortal Questions.* Cambridge University Press.

Nichols, S. forthcoming. "How Can Psychology Contribute to the Free Will Debate?" In J. Baer, J. Kaufman, & R. Baumeister (eds.) *Psychology and Free Will*, Oxford University Press.

Nisbett, R. and Cohen, D. 1996. *Culture of Honor.* Westview Press.

Peristiany J.G. and Pitt-Rivers J.A. 1992. *Honor and Grace in Anthropology.* Cambridge UP.

Pereboom, D. 2001. *Living Without Free Will.* Cambridge University Press.

Rawls, J. 1972. *A Theory of Justice*. Oxford University Press

Richerson, P., and Boyd, R. 2005. *Not by Genes Alone.* University of Chicago Press.

Smilansky, S. 2000. *Free Will and Illusion.* Oxford University Press.

Sommers, T. 2006. "The Two Faces of Revenge: Moral Responsibility and the Culture of Honor." (Paper presented at 2006 meeting Society for Philosophy and Psychology, and the 2006 meeting of the International Society of Research on Emotions (ISRE). Currently under review.

Sommers, T. 2007. "The Illusion of Freedom Evolves," in *Distributed Cognition and the Will*. Spurrett, D., Kincaid, H. Ross, D., Stephens. L. eds. MIT Press.

Sommers, T. (forthcoming). "More Work for Hard Incompatibilism." *Philosophy and Phenomenological Research.*

Smith, M. 1994. *The Moral Problem.* Blackwell.

Strawson, G. 1986. *Freedom and Belief.* Oxford: Clarendon Press.

Van Inwagen, P. 1983. *An Essay on Free Will*. Oxford: Clarenden.

Vargas, M. 2004. "Responsibility and the Aims of Theory: Strawson and Revisionism" Pacific Philosophical Quarterly 85:2 (2004): 218-241.

Vargas, M. 2005. "The Revisionist's Guide to Responsibility" *Philosophical Studies* 125:3 399-429

Williams, B. 1993. *Shame and Necessity*. University of California Press.

Wolf, S. 1987. "Sanity and the Metaphysics of Responsibility." in F. Schoeman (ed.), *Responsibility, Character and the Emotions: New Essays in Moral Psychology*. Cambridge: Cambridge University Press