Moral Responsibility: A Story, an Argument, and a Vision

Stephen J. Morse

University of Pennsylvania Carey Law School

Repository Citation
https://scholarship.law.upenn.edu/faculty_scholarship/886
Moral Responsibility:
A Story, an Argument, and a Vision
Stephen J. Morse

The concept of moral responsibility and associated concepts and practices, such as human agency, praise, and blame, play an undeniably important part in our lives. But what does it mean to be a morally responsible person—an agent that can fairly be praised and blamed, rewarded and punished, for one’s intentional actions? Many people look for a secure foundation to ground concepts as important to us as responsibility. Most notable among the seekers are theologians and secular metaphysicians, who try to justify responsibility with speculations and arguments about the divine or ultimate reason. I am unconvinced by these accounts, however, and believe that no ultimate, uncontroversial justification can be found. There is simply no way to know about the existence of God or the genuine ontology of the universe. What is more disquieting, if we were convinced that God or metaphysical moral reality existed, such a conclusion would not lead to uncontroversial concrete answers to the specific questions of morals and politics that vex human life. For example, even if we agree that in principle there are ontologically correct answers to every moral question, does this tell us clearly whether, say, the death penalty or abortion is ever justified? On epistemological grounds alone, then, if foundations exist, we will never know that we have reached rock bottom.

I will offer instead a story, an argument, and a vision about how moral responsibility is possible and why it is desirable. My account is an explicitly internal interpretation and defense of our present moral responsibility concepts and practices. Responsibility and its practices, like the concept of law and legal institutions, are products of human culture and can only be justified by the tools of justification that a particular culture provides. Although cultural products, including theories of knowledge and justification, may be temporally and geographically diverse, they are nonetheless as real features of our lives as gravity and death, and they can be rationally analyzed. To give an internal, constructivist account does not commit one to “anything goes” relativism or to unfettered pragmatism. We can assess the quality of the reasons that support moral responsibility, and I suggest that responsibility is both coherent and desirable. Indeed, our lives and society would be impoverished without it, even if it were possible to give it up, which I doubt.

I begin the story at the beginning. Once upon a time the universe began. I am agnostic about cosmology, but any secular theory will do. For convenience, let us assume that in the beginning was the “big bang,” as a result of which the universe could have been composed of matter or anti-matter. In the event, it turned out to be matter. Then a very great deal of physical lifting and hauling took place over a very long time. Now, this process was and is entirely governed by the physical laws of the universe. This thought is expressed nicely in Bernard Malamud’s novel, *Pictures of Fidelman.* A light bulb begins to speak in light to the protagonist, Fidelman, and gives him advice. Fidelman asks for guidance to follow the advice and the bulb says, “I will show you the way but I can’t go with you. Up to a point but not further if you know what I mean. A bulb is a bulb. Light I got but not feet. After all, this is the Universe, everything is laws.” All those laws operated upon all that matter to produce those features of the universe that have existed in the past, that exist now, and that will exist in the future. Among these features is our planet, Earth, on which those physical processes ultimately produced organic, biological life forms.

Those life forms evolved into wondrously diverse creatures, some of which, including ourselves, are “social” animals. To the best of our knowledge, the social life of most of these creatures, even highly complex forms, operates entirely, or almost entirely, “instinctually.” That is, primarily the genetic code provides these creatures with a repertoire of social
organization and behavior that is not self-consciously evaluated, adopted, or revised by those creatures. Human beings are not only social, however. Evolution has also endowed us with self-consciousness and the capacity for reason. “Instinct” may motivate our sociability, but compared to other creatures on this planet, it puts only the loosest limits on the forms that sociability takes. Among the species that inhabit our planet, only human beings act self-consciously for reasons and live in societies governed by behavior-guiding norms that are used in practical reasoning. We are the only creatures whose behavior is reason-responsive. This means that we are the only creatures to whom the questions, “Why did you do that?” and “How should we behave?” can properly be addressed and answered with reasons. To the best of our knowledge, only human beings kill each other as a result of disputes about the answers to such questions.

How could it be otherwise? Sociable, self-conscious creatures that reason have a lot of work to do if they are to live together at all successfully. They will need norms to guide their interactions. Many of these norms may simply solve ubiquitous coordination problems. But not all norms solve only these problems. For example, norms about when it is acceptable to kill another human being involve more than simple coordination problems. The norms may be temporally and geographically diverse and differential importance may attach to similar norms in different times and places, but norms there must be for creatures such as ourselves. It is nearly impossible to imagine a society that we could intelligibly consider human that existed without norms.

Norms in turn create in people mutual expectations of each other that vary in strength and importance. Indeed, we are probably the only creatures on Earth that consciously have normative expectations of each other and that make evaluative judgments when those expectations are satisfied or breached. Normative judgment appears to be an omnipresent and unique feature of human social existence. An interesting feature of human normative judgments, which have propositional content, is that they often motivate our emotional reactions. If someone pushes me and I fall hard, my emotional response will depend on why I was pushed. If it was done to save me from danger, I will feel grateful; if it was done to hurt me, I will feel angry, resentful, indignant. What emotions we feel in response to the satisfaction or breach of expectations may be largely culturally relative, but an emotional response to normative judgment seems to be another omnipresent aspect of human interaction. Once again, a society in which members did not have emotional responses to normative judgments would not be recognizably human.

In sum, creatures built like ourselves inevitably will live in groups that have norms, which in turn create expectations that generate emotional responses to their satisfaction or breach. It is difficult to imagine that these general aspects of human life would not exist unless human consciousness and the capacity for emotional response were radically altered. “Brave new world” hasn’t arrived nor has anyone yet given an ultimately convincing reason why we should hasten its arrival—although many have tried—so I propose for the nonce that we take ourselves largely as we find us.

So far this abstract story is largely descriptive, rooted in a loose, but not implausible, evolutionary biological account, and it does not commit anyone who believes it to any specific form of moral or political life. It commits one only to believing that among human beings there will inevitably be a moral and political life, that there will be norms, normative expectations, and reactions to the satisfaction or breach of those expectations. It applies equally well to any time and place within recorded history, and I suspect that it has been true since homo sapiens first used language to regulate social life. I freely confess that the story is a little bit “essentialist” about human nature, but this essence, although important, is not terribly limiting. The story is also a “determinist” or “universal causation” story, that is, an account that assumes that what happened was the outcome of a lawful set of physical processes operating on antecedent conditions.

At this point in the story, one might fairly ask why any defense—a normative task—of moral responsibility or any other cultural artifact is necessary if the deterministic or universal causation account is true. If, that is, what exists is the outcome of antecedent events and the laws of the universe operating on those events, then isn’t it true that no other outcome was
possible? If so, any evaluation or defense of "what is" is, itself, deterministically or causally inevitable and also needs no defense. Indeed, my students always ask just this question just about now. Well, in my story, the answer is, yes and no.

Yes, it is true that culture and everything else are the products of deterministic or causal processes, or something very close to them. But, no, it does not follow that cultural artifacts cannot and should not be normatively evaluated. Permit me to share how I attempt to demonstrate this conclusion to my students. I wordlessly move to the desk at the front of the class and then sit silently and as absolutely still as I can until the anxiety in the room builds to uncomfortable levels. I then ask the students what I was doing. They of course haven't a clue about why their professor behaves so bizarrely. The answer I give is this: "I am waiting for determinism to happen." They titter at the absurdity of the statement, and well they should. Although the winds, tides, bacteria, and most non-human species do "wait" for determinism to happen, human beings do not. We cannot help but consciously think, deliberate, reason, and guide our actions by reasons. We're stuck. It's just the way we are and there is no getting around it. Some unforeseen biophysical calamity or new technologies might change our nature. But short of such scary scenarios, there is no alternative but to deliberate and to act based on deliberation.

Even if our actions are determined, they make a difference. Determined actions can cause pleasure or pain, can create wealth or poverty, can be kind or cruel. We discuss, argue, fight, and even kill about morals and politics precisely because the moral and political regime in which we live makes an enormous difference to our well-being and flourishing. Although we may be determined creatures, we are not automatons. Determined deliberation and intentional action are distinguishable from reflexive bodily movements. We can't wait for determinism to happen. We determine what determinism dictates.

The tools we use to engage in such determined deliberation are given to us by our history and culture, but the human capacity for rationality and practical reason necessarily infects the process everywhere and always. How could anyone ever rationally persuade anybody else about anything, except by giving reasons? Unless we are satisfied to "convince" by force, all that we have is reason, however flawed it may be and however often it may lead us astray. The best we can do is to try to give good reasons and to yield to what seem to be better reasons when they appear. Although "reason" and "rationality" are not self-defining, uncontroversially-defined terms, it is almost impossible to conceive of a culture that will not include among its tools some criteria for what counts as good reason within the culture. Arguments that are logically sound—within defensible conventions of logic—and factually accurate—within defensible epistemological conventions—should yield only if they lead to a social and moral regime that there is better reason to reject. I propose to use ordinary logic and commonsense rationality—as they have been given to me and as they appear reasonable to accept—to continue my story. Anyone who objects needs to give me good reason not to and anyone who tries to do so has already accepted the most general claim being made.

It is impossible to deny that we hold each other morally accountable and have a rich series of practices that reflect this response. But what are we doing when we hold people morally responsible for their conduct? A defense of moral responsibility should begin with an interpretation of our practices that answers this question. I have previously suggested that human beings possess the innate capacity to respond emotionally to normative judgments. Following Sir Peter Strawson and more recent Strawsonians, especially Jay Wallace, I contend that holding people responsible is an expression of this capacity: it is the disposition to feel appropriate emotions when moral expectations are satisfied or breached and to express those emotions in the appropriate way. The content of the expectations, the appropriate emotions, and the appropriate expression of them may vary, but this is the meaning of the abstraction, "to hold someone morally responsible." In our culture, the common emotional responses to breach are resentment, indignation, and anger; the common emotional response to compliance is gratitude. Depending on the nature and magnitude of the expectation, the appropriate expressions of the reaction to breach can range from mild disapproval to painful punishment; the appropriate expressions of the reaction to compliance can range from mild approval to substantial rewards.
Moral Responsibility

The various ways in which the moral reactive emotions can be expressed must be normatively justified because they have the potential to create benefit or to cause harm. I say this not as an assertion of theoretical truth divorced from our practices, but as an expression of our internal practices, of what we require of each other morally. As a general rule, benefits and harms should not be awarded for “no good reason”; there should be good reason among creatures who self-consciously decide how they should live together. People should be praised and blamed, rewarded and punished, only if they deserve it, only if it is fair to respond in these ways to their conduct. To claim that reason should not guide the distribution of rewards and harms is to adopt an arational stance. But why should I or anyone else do this unless we are given good reason to do it?

Moral expectations give people good reason to behave properly. It is fair to hold people to such expectations only if they have the general capacity to grasp and be guided by reason. If they are incapable of understanding expectations or of using them in practical reasoning, it would be unjustified and unfair to praise and blame them for compliance and breach. Thus, we exempt people from moral responsibility under two general conditions that indicate that the agent could not grasp or be guided by reason. The first is cases in which the agent has not acted, that is, her movements did not meet the criteria for human action. Reflexes and other physically involuntary bodily movements are the prime examples. The second is cases in which the agent has acted, but the action is excused. Non-culpable ignorance, irrationality, and compulsion are classic examples.

Under either condition—no action or excused action—we do not think it is justified to feel and express what would otherwise be the appropriate emotion for compliance or breach. For example, if a driver has an unforeseeable seizure and blackouts, during which the car causes injury, we would not justifiably be angry at and sanction the driver for the moral failure to take care while operating a dangerous machine. Or, under the same circumstances, if the car plows harmlessly into a haystack, saving a passenger from an otherwise imminent but unforeseeable crash with an approaching vehicle, the unconscious driver does not deserve thanks for saving the passenger’s life.

Although I have described the practice of holding people morally responsible as a predisposition to an emotional reaction and an expression of that reaction, holding responsible is not a purely emotional response. A moral reactive emotion and its expression are appropriate only if a moral expectation was complied with or breached and the agent was capable of being guided by reason. The criteria for moral expectations, action, and reason may be variable, but they are always propositional and have truth value. Moreover, our practice of holding people morally responsible is not simply an instrumental behavioral disposition to emit responses that are meant to increase or decrease the probability that particular behavior will occur in the future. Holding people morally responsible expresses a normative, retrospective evaluation of an agent’s conduct, independent of whether that evaluation has positive or negative consequences.

In sum, the disposition to feel appropriate emotions when moral norms are breached or satisfied and the consequent expressions of normative judgment affirm and enforce the inevitable, necessary norms of behavior that exist in human society and that give people reasons to act. After all, the elements of nature, such as the winds and the tide, and non-human species do not need norms or rules of conduct to tell them how to behave, nor do we feel moral emotions towards them, unless we anthropomorphize them. I assume that expressing normative judgments and affirming and enforcing norms could be accomplished by means other than holding people morally responsible, although I also assume that this would be a difficult trick to pull off. Consistent with this assumption, however, the defense of responsibility will still require a story about why it is desirable—why, that is, human beings should rationally desire to live in societies that hold people responsible for their intentional conduct.

Should our concepts and practices of moral responsibility be maintained? The first line of attack is the familiar one, alluded to above, rooted in anxieties about determinism or universal causation. I previously suggested that various interpersonal, social concepts and practices were real and could be rationally interpreted even if determinism or universal
causation were true. Determinist critics of responsibility—so called “hard determinists” or “incompatibilists”—can concede these points, but they argue nonetheless that responsibility isn’t really real; it’s just a made-up, as-if fantasy.” Thus, there is “nothing” desirable to maintain, and it is entirely unfair to maintain practices dependent on such fantasies. After all, if all events in the universe are determined or caused products of antecedent events and laws over which living human beings have no control, how can we possibly be responsible for anything? Such critics believe that unless we have contra-causal freedom (or some such thing)—the ability to act unconstrained by the causal, deterministic processes of the universe—then “real” responsibility is not possible. And neither, it must be conceded if this is correct, are “real” justice and fairness possible, which presuppose rational, responsible responses of rational creatures to each other. If correct, this assertion of the incompatibility of determinism and responsibility (and justice) is powerful, because the metaphysics of contra-causal freedom are preposterous, and in Strawson’s word, “panicky.”

My account of responsibility is compatible with the truth of determinism or universal causation, however. Even if our behavior is determined, we are determined to have a susceptibility to feel and express emotions in response to compliance with or breach of moral expectations, and some human beings are capable of being guided by reason and others, such as small children and some people with severe mental disorder, are not. Responsibility does not require god-like causal powers. It requires only that mature human beings have the general capacity to grasp and be guided by good reason, a capacity we firmly believe we have. It is fair, we believe, to hold accountable another person who is capable of reason, a person who can use moral expectations as premises in practical reasoning. A person who suffers an unforeseen seizure may be causally responsible for harms that result, but the person does not deserve, in the fullest sense, our condemnation and sanction. Such a response would be unfair. If we didn’t believe that mature humans were capable of grasping and being influenced and guided by reason, we wouldn’t be reading or writing articles like this one in the hope of learning or being guided ourselves or of persuading or guiding others. Few people read or write academic articles solely for recreation or even advancement.

If contra-causal freedom is necessary for “real” responsibility—an assertion that flows from an external point of view and critique of responsibility”—then “real” responsibility does not exist, but I deny that it is necessary. If we all had god-like powers, then I suppose that there would be no question about responsibility. We are not gods, of course, but in a sense, we are god-like in that we must create and do have the capacity to revise our concepts and practices, even if we lack contra-causal freedom. And those creations are “real enough” for me, even if they lack the alleged metaphysical pedigree of contra-causal freedom.

Yes, we create and revise our responsibility concepts and practices over and over again as we go along, just as we do with the concept of justice and its practices. But what we construct is real, it expresses our nature, and it has the all-too concrete potential to benefit or harm us, to give our lives more or less meaning, or to extinguish those lives. The norms of responsibility are like the rules of law, the demands of justice, and the institutions of government: all have the enormous potential for good and ill, and they can be defended best by the good reason that we can give for them, rather than by a foundational metaphysics. Practical and theoretical reason is all that we have to devise and defend our concepts and institutions. Given the biological limits of our beings, what more could we realistically and reasonably want?

The retrospective, evaluative concept of moral responsibility that we now possess and employ may be real enough, but is it desirable? The second line of attack on responsibility suggests that it is an irrational concept. If one function of moral responsibility is to guide interpersonal life, then perhaps, as many have suggested, we could replace it, for example, with a purely prospective, purely consequentialist scheme. But in addition to being inadequate interpretations of current concepts and practices (Wallace 54-6), abandoning our current concept and practices would not be rational because it would impoverish our lives.

Retrospective, evaluative moral responsibility is crucial to our sense of ourselves as persons, as objects of dignity and respect, and it coheres with other moral notions of supreme importance, such as desert, fairness, and justice, to which we are firmly committed. What
Moral Responsibility

makes us distinctively human and what gives all these concepts full meaning is our capacity for reason, which in turn makes us capable of genuine normative evaluation and appropriate objects of such evaluation. Other sentient creatures can suffer and deserve to be treated without unnecessary pain, but we are the only creatures capable of leading fully moral lives. This is why, I believe, we think of ourselves as occupying a unique position among the life forms on earth and it is a foundational part of what gives our lives dignity, meaning, and worth. To give up this sense of ourselves would abandon our humanity.

Responsibility is desirable also because it contributes to the creation and maintenance of moral communities. Personhood, desert, and responsibility are all moral notions, inextricably intertwined in our moral lives. When we express the reactive emotions through appropriate practices, we affirm and deepen our commitment to common moral obligations that bind us together. To diminish or to abandon moral responsibility would be to weaken those ties and the communities that nurture us. Creatures capable of grasping and being guided by good reason deserve our moral praise when they comply with our moral expectations, and they deserve our blame when they transgress. To make this claim does not commit one to any concrete scheme of expectations, rewards, and sanctions. It simply commits one to the recognition that we are the only creatures capable of being guided by reason and that this is an inevitable, crucial feature of our social lives.

Many people object to holding people morally responsible, for fear that this practice readily produces moralistic, punitive, and cruel responses, at least when expectations are breached. It is of course true that when human beings are angry because they have been harmed, they can become incredibly self-righteous and hideously cruel. Holding people morally accountable can be used to satisfy such unworthy responses, but this is not inherent in the practice. Virtually any human practice can be turned to vicious use, but nothing in my story compels moral responsibility to be applied harshly. My account is consistent with a tender, forgiving, or conversely, a tough, strict set of norms and practices, and neither need be cruel. It presupposes only that it is fair to hold accountable beings that reason and that are capable, as most adults demonstrably are, of being guided by good reason.

Another conception of human life is, I suppose, possible. For example, we could treat each other like bacteria. I refer to this as the good bacteria/bad bacteria part of the story. Some bacteria that inhabit our gastrointestinal system, our gut, are crucial to the smooth operation of the system. They are the good bacteria. We try to enhance their survival and do nothing to inhibit their growth. On occasion, alas, our guts are invaded by bacteria that interfere with the proper operation of the system, causing various unseemly ailments, and in extreme cases, death. These are the bad bacteria. We spend a fair amount of effort trying to prevent these critters from entering our gut in sufficient numbers to overwhelm the body's natural defenses, and if the natural defenses fail, we try with various techniques, such as antibiotics, to kill the offensive, bad bacteria. Now, despite the potential of various bacteria to confer benefits and harms, as the case may be, and despite our consequential, substantial efforts to deal with these bacteria, no one holds either kind of bacteria responsible for smooth or rocky gastrointestinal functioning, and we wouldn't dream of praising or blaming bacteria. We treat bacteria purely as objects, and never as subjects, as agents.

We could, by analogy, simply treat each other like bacteria, as potentially beneficial or harmful objects, and act accordingly. This conception of people would support a purely predictive and preventive scheme of social organization in which the emotional and societal response to the organism could be entirely independent of the moral goodness or badness of the person's conduct. We don't at present have the emotional repertoire or the predictive and therapeutic technology to institute this vision very precisely or effectively, but this is a technquibble. In principle, it is a possible form of social organization. Indeed, in some senses we might all be “safer,” and, to some, social life might appear more rational if the show ran along these lines. But this is a show that I think virtually any of us would happily miss, precisely because we are importantly different from bacteria. They are not moral creatures.

At this stage in the story, I am appealing as much to a vision as to a logical argument. Why would a rational agent wish to give up a conception of self as a moral creature? No one would do this, for example, just because determinism were true. A rational agent would do it
only if it promoted some vision of human flourishing the agent was willing to defend. If you are still convinced that moral responsibility contributes little and should be abandoned, try the following thought experiments. Imagine, first, that someone intentionally injures you, knowing that you are incapable of retaliating, because the person enjoys your pain and the feeling of dominance that cruelty produces. You of course react not only by feeling pain, but surely with anger, too. Your physical pain is no worse than it would be if an accident of nature had caused it, but your entire reaction is quite different. Do you really want to treat the bully as if she were just "bad bacteria"? Should your response be simply that which most decreases the likelihood that the bully would strike again? Do you think you (and others) could learn not to feel indignation, resentment, and anger, as opposed to regret? Would we all be better off if we no longer had these reactions and expressed them to each other? Would life then be more rational? Better? Would human flourishing increase?

Now imagine that you have unjustifiably injured another person. Probably, in addition to whatever "positive" emotion is created, you will also feel guilt or shame or some other critical emotion. But how should you feel about yourself? If you promise not to do it again and really, really mean it, should that be the end of the matter? Of course, unpleasant emotions like guilt or shame may decrease the likelihood that you will cause such injury again, but should that be the only point of these emotions? How should the victim feel about you? Is resentment and condemnation just irrational or silly, especially if you promise to be a good bacterium from now on? Finally, imagine that we all learned our lesson and ceased to have these irrational responses to violations of moral expectations. What would society be like? How would we feel about ourselves and each other? What would be the role of community or the arts in such a society?

My story has neither a happy nor a sad ending. We are in the midst of the story and have no idea whether or how it will end. Human social life will surely end only when human beings cease to exist as a species in its current form. As long as recognizable humans exist, however, they will create meaning and norms from the cultural tools and concepts available to them. And history and anthropology teach us that incredible diversity is possible. We cannot even predict that there will be progress. Progress is itself a socially constructed moral notion, subject to diverse interpretations.

If—and this is a big "if"—we insist on trying to subject our concepts and practices to reason, our capacity for reason surely sets some limits on what concepts and tools a society can adopt that will remain relatively stable. Now, of course, one can claim that reason is not the road to moral or any other kind of progress and that it is a mistake to try to subject our lives to the dictates of good reason. But the only reason anyone should accept such a claim is because the claimant gives good reason to abandon reason. Force and the threat of force are, of course, good instrumental reasons to yield to a stronger opponent—infracultural species do this all the time—but fear is not, in itself, a good moral reason for action.

No culture exists or changes without individual agents to perpetuate, criticize and modify it. Surely no culture changes unless individuals believe that it should and then act on that belief. We are all the creators as well as the consumers of our culture. As such, we are all storytellers as well as characters in the story. We have been born in a time and place in which we have been taught to trust reason, to hope for rationality and objectivity. Although we should not make reason a fetish, we ignore it at our peril. Because our norms and rules must be expressed in language, there will always be ambiguity and room for interpretation and argumentation. There will never be absolute agreement on what reason demands, even among people who share the same ends and the same vision of human flourishing. The best we can do is to give each other the best reasons we can for why we should live together one way rather than another, why we should prefer one set of norms, rules and institutions to others, and try to listen to others' reasons as openly as possible.

Whether or not taking moral responsibility seriously is a given of interpersonal life, it is fair and fundamentally enhances personhood, dignity and respect. It also facilitates the formation and maintenance of moral communities. These goals are so important that it is hard to imagine giving them up, at least not until someone gives us good reason that we should. And, if the relation between responsibility and these other ends is as close as it appears to be.
Moral Responsibility

it is hard to imagine what reasons would persuade us to abandon responsibility. If you wish to give it up, assuming that it is possible to do this, what social concepts and practices do you propose to adopt instead to respond to the good and evil we do to each other, to the intentional acts of kindness and cruelty of interpersonal life? What kind of society will it be? Tell us a story that gives us good reason to hear it again.

Notes


6 I borrow this felicitous term from Wallace.

7 For an interesting explication of the contrary pulls of internal and external points of view about responsibility (and everything else), see Thomas Nagel, *The View from Nowhere* (New York: Oxford University Press, 1986), esp. 120-126.

8 See, for example, Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (New York: Oxford University Press, 1984), ch. 7.