

1 Moral Responsibility

T. S. Eliot (1943, 37) speaks of “what was believed in as the most reliable—and therefore the fittest for renunciation.” Eliot could have been describing moral responsibility. It is believed in fervently. As Cicero (44 BCE/1923, 119) noted, philosophers are willing to entertain almost any hypothesis: “There is nothing so absurd but some philosopher has said it.” But even philosophers find it difficult to contemplate the renunciation of moral responsibility. Peter van Inwagen is typical: “I have listened to philosophers who deny the existence of moral responsibility. I cannot take them seriously” (1983, 207). And Peter Strawson insisted that “we cannot take seriously” (1962, 74) the rejection of moral responsibility and the radical changes that it would involve.

Commitment to moral responsibility is based in visceral emotional reactions and locked in place by a far-reaching theoretical system. But the moral responsibility belief system is fighting a running retreat against scientific research that renders this system less and less plausible. The purpose of this book is to show that the key arguments for moral responsibility fail, that moral responsibility is fundamentally inconsistent with our naturalistic world view, that we would be better off if we rejected moral responsibility, and that the abolition of moral responsibility is a genuine possibility; in short, that belief in moral responsibility is a widely held doctrine that is indeed “fittest for renunciation.”

What the Moral Responsibility Debate Is About

The dispute over moral responsibility is an old one, with many twists and turns. Some of those twists have involved disputes over exactly what is involved in saying that someone is *morally responsible*. The moral

responsibility that is my target is the moral responsibility that justifies special reward and punishment. Moral responsibility provides the moral justification for singling an individual out for condemnation or commendation, praise or blame, reward or punishment. If Susan justly deserves punishment, she must be morally responsible for the wrong she committed. Various philosophers offer a variety of grounds for Susan's moral responsibility: Chisholm and Campbell would say she could have done otherwise, Frankfurt that she reflectively approved of her own will, Fischer that she could exercise guidance control, Dennett that she passed a basic competence threshold. But whatever the conditions required for moral responsibility, it is meeting those conditions that makes punishment (and reward, blame, and praise) fair and just.

The purpose of this book is the abolition of the moral responsibility system, root and branch: we should *never* hold *anyone* morally responsible. It is essential to be clear about what that does not involve. As will be argued in subsequent chapters, it does not involve the rejection of all moral evaluations: Joe may do something that is morally wrong, Joe's immoral behavior may stem from his deeply flawed character, and it is important to recognize and examine those wrongs and flaws, but Joe does not deserve blame or punishment. And it may be useful to blame or punish Joe (though I very much doubt it), but Joe does not justly deserve such blame or punishment. As I use the phrase in this book, "moral responsibility" is the essential (necessary, if not sufficient) condition for justified blame and punishment. Michael McKenna states that "what most everyone is hunting for . . . is the sort of moral responsibility that is desert entailing, the kind that makes blaming and punishing as well as praising and rewarding justified" (2009, 12). What McKenna describes is precisely what I am hunting for as well; the difference is that rather than trying to preserve it and justify it, my goal is to kill it and drive a stake through its heart. But the present point is that when I take aim at moral responsibility, what McKenna describes is my target. That is the dominant way in which "moral responsibility" is understood, both in philosophical and popular use. For example, it is the concept of moral responsibility adopted by Galen Strawson: "responsibility and desert of such a kind that it can exist if and only if punishment and reward can be fair or just without having any pragmatic justification, or indeed any justification that appeals to the notion of distributive justice" (2002, 452). And it is the account given by Randolph Clarke:

If any agent is truly responsible . . . that fact provides us with a specific type of justification for responding in various ways to that agent, with reactive attitudes of certain sorts, with praise or blame, with finite rewards or punishments. To be a morally responsible human agent is to be truly deserving of these sorts of responses, and deserving in a way that no agent is who is not morally responsible. This type of desert has a specific scope and force—one that distinguishes the justification for holding someone responsible from, say, the fairness of a grade given for a performance or any justification provided by consequences. (2005, 21)

I am not claiming proprietary rights to how “moral responsibility” is used; rather, this is only a statement of how it will be used in this book.

So exactly what is this deep philosophical controversy over moral responsibility really about? First, it is not about the conditions for moral responsibility in our current system of holding people morally responsible. That is an important question, certainly, and lawyers and justices and legal scholars have joined philosophers in that debate. This internal debate concerning the proper details of our system of moral responsibility and justice includes difficult questions. Who is competent? What is the age of responsibility? Do addictions destroy moral responsibility? What excuses are legitimate in the moral responsibility system? What is the appropriate legal standard for insanity, and can one be insane and still be morally responsible? A brief perusal of the legal literature on insanity shows how difficult and controversial such questions can be. Those are questions within the system of moral responsibility—difficult questions that have occupied many astute and insightful thinkers. But those are not the questions being examined in this book.

My concern is not with the details of our moral responsibility system, but with the system itself. When philosophers such as C. A. Campbell, Robert Kane, John Martin Fischer, Alfred Mele, Derk Pereboom, Michael McKenna, Saul Smilansky, Susan Wolf, Daniel Dennett, and Randolph Clarke wrestle with the question of moral responsibility, their basic concern is not the details of when our given moral responsibility system justifies and excuses from moral responsibility. Their question is more fundamental: is it ever morally justifiable to hold anyone morally responsible? Rather than the internal details of our system, they are struggling with the question of whether our system of holding people morally responsible is itself morally justified. Is our overall system—or any system—of moral responsibility and “just punishment” really fair? Can our system, or any moral responsibility system, withstand close scrutiny? Can any system of just

punishment be morally and rationally justified? That is, they are wrestling with the basic external question of moral responsibility.¹

It is important to distinguish the internal debates from this fundamental external debate, because too often the lines get crossed. If someone argues, for example, that miscreants who suffered an abusive childhood are morally responsible and justly deserve punishment because our system of moral responsibility does not recognize a harsh childhood as a legitimate exemption from moral responsibility, that's fine; someone else may argue that given the precedents within our system, consistency requires extending excusing conditions to such unfortunate individuals. That is a fascinating internal argument, and it may be pursued with great vigor and insight and resourcefulness. But if someone takes that argument and moves it to the external controversy, then the argument begs the question by assuming the very system of moral responsibility that is in external dispute. That is, if someone argues that murderers who suffer abusive childhoods are not excused from moral responsibility and justified punishment (in our system) and that therefore, at least some people *are* morally responsible, and moral responsibility itself is (externally) justified, then such an argument fails: it assumes what (at this external level) it is supposed to be proving. This distinction means that data on how "moral responsibility" is commonly used—though certainly fascinating—cannot settle the basic external question concerning moral responsibility. I doubt that there is an internal ordinary language usage of "moral responsibility" that is consistent,² but there is no doubt that within our culture, people do make claims and ascriptions of moral responsibility. Within the moral responsibility system are frequent assertions of moral responsibility; the external question is whether that system is justified. "Moral responsibility" is in common use and typically understood within that system, but that usage no more justifies the system itself than the common use and understanding of "witch" justified the brutal system that consistently identified and executed witches.

Second, the external debate over moral responsibility is not a debate over whether anyone ever does anything morally wrong or morally right. If no one ever did anything morally right or wrong, the question of moral responsibility would be reduced to a very abstract intellectual exercise. The question is not whether anyone ever does wrong, but whether those wrongdoers justly deserve punishment, whether those who do wrong (and

those who do right) are morally responsible for their vicious or virtuous behavior. (Some claim that without moral responsibility there could be no judgments of right or wrong, virtue or vice; that claim will be considered in due course, but I trust that it does not reveal too much of the plot to say that I believe the claim to be false.) Furthermore—and this argument will raise heated objections, which must be considered in later chapters—the basic external debate over moral responsibility is not a question of whether people can have vicious or virtuous characters. The claim here is that they can, but that they do not deserve praise or blame for their characters, or for the behavior which flows from their character traits. No one is morally responsible for being bad or behaving badly—but this does not mean that no one has a character with profound moral flaws.

Third, the debate is not about the efficacy of moral responsibility practices. If someone asserts that the moral responsibility system works well in preventing crime and improving character, there would remain the more basic question: yes, but is it really just? If we could keep a wonderful system of law and order by sacrificing one person chosen at random every year, that might be a tempting tradeoff; indeed, if God offered us such a system, we might well sign on (especially when we consider that otherwise more people will be wrongly killed each year). But we would still have the question: yes, but is it really just? Did the punished person genuinely deserve punishment? Perhaps the larger benefits of this system would outweigh the injustice done to the innocent person who is sacrificed, but that would not change the fact that the innocent person is punished unjustly. The practice of moral responsibility is not an effective way of producing either a safer society or better behaved individuals. To the contrary (it will be argued): moral responsibility blocks implementation of much better systems and causes enormous suffering. But even if the practice of moral responsibility were effective in making a better society, this effectiveness would not prove that the system is just. The point here is only that (unless one is a narrowly doctrinaire utilitarian) it is one thing to determine that a system is efficient—and quite a different process to decide whether it is just.

Accountability and Moral Responsibility

Moral responsibility views that shift the focus away from morally justified punishment and reward fail to capture our basic notion of moral

responsibility. Some claim that moral responsibility is about applying punishment and reward where they will produce the greatest social benefit. Others claim that moral responsibility involves only the making of moral judgments: when we say that Beverly did something morally wrong, that necessarily involves the judgment that she is morally responsible for her bad behavior. Several of those competing standards for moral responsibility will be examined and critiqued in the course of this book. But one such position is quite elegant and will get immediate attention: the view that moral responsibility is about *accountability*. If you are morally responsible for an act, then (on the accountability model) it is legitimate to require you to give an account or justification for your act. When we say that Cassandra is morally responsible for an act, we mean that she is accountable for that act: she must be capable of giving an account of why she acted.

Accountability, however, fails as a standard for moral responsibility. In the first place, people often give an account of why they did something—a sincere and honest account—that is completely mistaken. There are many social psychology experiments in which people act under the influence of factors outside of their awareness, factors that they would vigorously deny having influenced their behavior. In one famous experiment (Isen and Levin 1972), people who found a dime in a phone booth almost all stopped to help a stranger who had “dropped” a set of papers, and those who did not find a dime seldom helped. Yet the helpers were blissfully unaware—and would certainly deny—that finding a dime could and did strongly influence their behavior: it would not be part of any account they would give of their helping behavior. Even more striking are the experiments in which part of the brain is externally stimulated, causing people to respond, but they give an account (in terms of their own motives and reasons) for why they “acted” as they did. José Delgado, a neuropsychologist, discovered that electrically stimulating a specific region of a patient’s brain caused the patient to turn his head from side to side, as if looking for something. When Delgado asked his patient to “give an account” of his behavior, “the patient considered the evoked activity spontaneous and always offered a reasonable explanation for it” (Delgado 1969, 115).

Even in cases in which an individual gives a correct account of why she acted badly (or well), that does not establish that she is morally responsible for that act. Suppose that Ann makes a bad decision and commits a morally

bad act because she decides hastily and fails to consider important moral factors that (had they entered into her deliberations) would have led her to a better choice. If we ask Ann for an account of her bad act, she might accurately report that she chose badly because she is a “cognitive miser” (Cacioppo and Petty 1982) who can deliberate, but who has never developed the capacity for sustained deliberation; indeed, she might regard her cognitive impetuosity as a moral flaw. But is Ann *morally responsible* for her cognitive shortcomings or for the flawed choices that stem from those limitations? Those are questions that still remain after Ann has given an account of her flawed behavior. Thus the capacity to give an account of one’s acts is not the same as being morally responsible for those acts.

This view—to be morally responsible is to be subject to the demand to give an account—is appealing because it associates moral responsibility with the special powers of rational account-giving capacity: powers often treated as transcending the histories that shaped us and the limited rational powers we actually enjoy. As tempting as such a view is, as long as we restrain rationality within naturalistic limits, it cannot justify moral responsibility—or so I argue in the forthcoming chapters. Thus if one insists on redefining “moral responsibility” in terms of “accountability,” the original question of moral responsibility will still be there after the accountability question is settled: Ann is accountable for (can give an account of) her behavior, but is she morally responsible (justly deserving of blame) for what she did?

Whatever the merits or faults of other views of moral responsibility, this book focuses on what I take to be the core concept of moral responsibility: moral responsibility is what justifies blame and praise, punishment and reward; moral responsibility is the basic condition for giving and claiming both positive and negative just deserts.

The Deep Belief in Moral Responsibility

My goal in this book is to show that claims and ascriptions of moral responsibility (in the robust sense specified previously) cannot be justified, that there are strong arguments to show that—absent miracles—the system of moral responsibility and “just deserts” is fundamentally unfair, and that we will be better off when belief in moral responsibility is utterly eliminated. But belief in moral responsibility is so deep and pervasive—among

philosophers and nonphilosophers alike—that it is necessary to pause for a moment and insist that my goal truly is the complete denial and rejection of moral responsibility. It is necessary to insist on that denial because many people (including many philosophers) find it difficult to imagine that anyone genuinely denies moral responsibility.

If we examine why so many regard moral responsibility as immune from serious challenge, then the basis of that belief does not seem nearly so solid and attacking moral responsibility may appear less quixotic. Moral responsibility has many dedicated defenders, but few of them would claim that compelling reasoned arguments have placed moral responsibility beyond challenge. Many philosophers believe that there are good arguments to support moral responsibility, but they do not regard the arguments as so conclusive that they are beyond doubt. Thus the unshakable certainty of many philosophers concerning moral responsibility must have some emotional source independent of rational argument. Perhaps that deep and widespread emotional commitment to moral responsibility is positive and defensible, but we are sadly familiar with many deep emotional commitments—to racism, sexism, jingoism, xenophobia—that examination reveals to be harmful and irrational. Of course some of our deep emotional commitments—our love for our children, for example—still look worthwhile after close scrutiny. But when we recognize that moral responsibility is rooted in emotions rather than reason, it should be less difficult to take seriously the possibility of rational philosophical challenge to the system of moral responsibility. At that point, we can carefully examine the actual arguments for moral responsibility, as well as the case against moral responsibility, and judge them on their merits.

The philosophical literature is replete with a wide variety of very sophisticated arguments in favor of moral responsibility, ranging from exotic varieties of libertarian speculation to mundane claims of pragmatic benefit. But those arguments do not purport to discover moral responsibility; rather, they are put forward to justify a visceral and universal emotional reaction: the basic retributive impulse, the deep desire to strike back when we are harmed. Legal philosopher and scholar Michael S. Moore describes this retributive desire quite clearly:

Of course Dostoyevsky's nobleman [who has his dogs kill a small child in front of the child's mother] should suffer for his gratuitous and unjustified perpetration of a terrible wrong to both his young serf and that youth's mother. As even the gentle

Aloysha murmurs in Dostoyevsky's novel, in answer to the question of what you do with the nobleman: you shoot him. You inflict such punishment even though no other good will be achieved thereby, but simply because the nobleman deserves it. The only general principle that makes sense of the mass of particular judgments like that of Aloysha is the retributive principle that culpable wrongdoers must be punished. This, by my lights, is enough to justify retributivism. (1997, 188)

Whether Jew or Gentile, courtly or common, Elizabethan or contemporary, the words of Shakespeare's Shylock resonate: "I am a Jew. Hath not a Jew eyes? hath not a Jew hands, organs, dimensions, senses, affections, passions? fed with the same food, hurt with the same weapons, subject to the same diseases, heated by the same means, warmed and cooled by the same winter and summer as a Christian is?—if you prick us do we not bleed? if you tickle us do we not laugh? if you poison us do we not die? and if you wrong us shall we not revenge?" (Shakespeare 1596–1598/1993, 3.1.58–68). Many philosophers seem to have a similar visceral commitment to retributive just deserts. Robert C. Solomon states: "Sometimes vengeance is wholly called for, even obligatory, and revenge is both legitimate and justified. Sometimes it is not, notably when one is mistaken about the offender or the offense. But to seek vengeance for a grievous wrong, to revenge oneself against evil—that seems to lie at the very foundation of our sense of justice, indeed, of our very sense of ourselves, our dignity, and our sense of right and wrong" (2004, 37).

Even Daniel Dennett, who has developed an incredible variety of sophisticated arguments to support moral responsibility, has recently rested his case for moral responsibility on our deep retributive desires:

We ought to admit, up front, that one of our strongest unspoken motivations for upholding something close to the traditional concept of free will is our desire to see the world's villains "get what they deserve." And surely they *do* deserve our condemnation, our criticism, and—when we have a sound system of laws in place—punishment. A world without punishment is not a world any of us would want to live in. (2008, 258)

Among contemporary philosophers, Peter French is the most outspoken in celebrating vengeance as a virtue, and he is brutally frank concerning the emotional roots of this view: "Personal and vicarious moral anger can be and ought to be placated by hostile responsive action taken against its cause. Wrongful actions require hostile retribution. That, despite its seeming lack of fit with the body of moral principles upheld in our culture,

is actually one of the primary foundations of morality. It is a foundation that is settled in passions, attitudes, emotions, and sentiments, not in reason" (2001, 97). Thus the source of fervent belief in moral responsibility is feelings: the powerful feeling that those who do wrong and cause harm should suffer. That feeling is rooted in an even deeper feeling: when we are harmed, we should strike back.

Perhaps those feelings can, ultimately, be justified by rational argument. I think not, and most philosophers think so; the immediate point is not whether such rational justification can be provided, but that the profound and common belief in moral responsibility is not the product of rational deliberation. Thus, regardless of whether the arguments to justify moral responsibility work, we should not be surprised to find widespread strong belief in moral responsibility. But that almost universal belief cannot itself be used to justify moral responsibility, any more than widespread belief in the existence of God or the justice of subordinating women can justify those beliefs. The strong and common feeling of male superiority was no justification of that view; even if one imagines that the widespread masculine sentiment in favor of female subordination is evidence that it once had some survival value, it is clear that what may have been useful at one stage of development is now maladaptive. The same may be true of belief in moral responsibility: even if one grants that at earlier stages of development, retributive practices were of some benefit (in terms of either group or individual selective pressures), it may be that in our present state they are maladaptive (just as human aggressive tendencies have become severely problematic in an era of handguns, not to mention nuclear weapons), as well as being morally and rationally unjustified and unfair.

When we examine the origins and nature of our retributive emotions, we may conclude that those emotions are not as attractive and virtuous as we had once imagined. Their roots are in a strikeback response that attacks whatever is near. When rats are placed in a cage with an electrified floor and then shocked, they attack one another. When a rat is hurt, its immediate desire is to strike back at *something*: its assailant, or an innocent bystander, or a gnawing post if nothing else is around. Rats that are shocked, but that can vent their rage against another rat (or a gnawing post), suffer fewer problems than rats with nothing to attack: they have less increase in adrenal hormone and blood pressure levels and develop

fewer stomach ulcers (Virgin and Sapolsky 1997; Barash 2005). In monkey colonies, a subordinate who suffers an attack from a higher-ranking monkey typically seeks out an individual lower in the hierarchy for attack (Kawamura 1967). Veterinarians warn that a gentle but territorial pet cat may become disturbed by the smell or sound of another cat in the vicinity, and if unable to attack the “intruder,” it may redirect its aggression by attacking a family member.

We might imagine ourselves well beyond such primitive reactions, but we understand perfectly when Curly hits Moe, and Moe then strikes Larry. A cartoon shows a boss berating a subordinate man, the man coming home to yell at his wife, the wife reprimanding the child, who then kicks the unoffending dog; a little funny and a lot depressing, but we need no explanation to see the sad humor. When layoffs and economic stresses increase, we reliably expect a corresponding increase in abuse of spouses and children. When the United States was hurting from the September 11, 2001, attacks, the assailants were either dead or elusive; Iraq, which had nothing to do with the attack, bore the brunt of U.S. strikeback anger. A nineteenth-century British jury, in the farmlands of Devonshire, once returned the following note when it found a young man guilty of stealing hay: “We don’t think the prisoner done it, but there’s been a lot taken hereabouts by someone” (Brown 1899, 513). And so *someone* has to suffer for it: the thief if handy, but an innocent if no one else is available. Unscrupulous prosecutors have long realized that if the case against a defendant is weak, then grisly photos of the crime scene and the murder victim and graphic descriptions of the brutal murder can be an effective substitute for the missing evidence: when jurors are outraged by a criminal act, their desire to strike back at somebody can easily overwhelm their concern about whether the defendant is the appropriate target of their wrath.

Analyze This is a film comedy about a mob boss who seeks psychological treatment. The mob boss, Giotti, is furious with a rival mobster who had attempted to kill him, and Giotti’s psychologist is offering counsel on anger management: “You know what I do when I’m angry? I hit a pillow. Just hit the pillow, see how you feel.” Whereupon Giotti pulls out a pistol and fires several slugs into the pillow. When the psychologist regains sufficient composure to ask “Feel better?” the mobster responds, “Yeah, I do.”

He does indeed feel better, and there's a good reason for that: when someone (human, chimp, or rat) is attacked or threatened or harmed, they experience "subordination stress": they suffer from chronically overactive stress responses. But if the attacked individual is able to attack someone else, his hormonal levels are then reduced and the stress is eliminated (Virgin and Sapolsky 1997).

How did this unattractive and inefficient emotional reaction become so deeply entrenched? Our ancestors who struck back—or lashed out at someone—when attacked or wronged were more likely to survive and flourish than those who "turned the other cheek" or reacted passively. As David Barash notes, if my attacker is too strong for a successful counterattack, by attacking someone else I serve notice that I am still someone to fear, someone who cannot be attacked (at least by subordinates) without serious consequences: "Evolution would most likely reward victims who—even if unable to retaliate against the actual perpetrator—conspicuously 'take it out' on someone else" (2005, 4). The reciprocal tit-for-tat approach—particularly the punishment of harms—was favored because it worked—not because it worked particularly well (it didn't and doesn't). Retributive impulses are often directed against innocent family members and scapegoats rather than wrongdoers. Even when aimed at an offending target, punitive or retributive responses are grossly inefficient, as behavioral research has long since established; in addition, the side effects (such as blocking inquiry into deeper systemic causes) result in even greater harm. Still, in the many millennia before behavioral science, the strikeback response—crude as it is—was the only available tool for social control and a means of protecting status, and it was better than nothing. Because it sometimes produced positive effects, it became biologically, socially, and psychologically entrenched.

The basic retributive emotion runs deep—much deeper than the history of our species. Belief in moral responsibility is more recent. If we look at early legal codes, moral responsibility is not an important element: an eye for an eye, a tooth for a tooth, and there is little or no distinction between causing harm purposefully or inadvertently. Questions of moral responsibility do not arise: God commands this punishment, and that settles it. No further justification is required, and it would be wicked to ask for one: who are you to question God? We have reports late into the medieval period of animals being executed for having caused a death, and questions

of whether the offending animal is morally responsible or justly deserving of punishment are not an issue. But—some exceptions notwithstanding—over the centuries, distinctions were made between wrongs committed purposefully and wrongs that occurred by accident and eventually between wrongs done by competent persons and wrongs committed by the demented or those who have not reached the “age of responsibility” (though in the United States an eight-year-old child was recently charged with murder; perhaps we have not come so far as we might have imagined). And as the system grew more sophisticated, eventually the most basic question emerged: what are the grounds for holding people morally responsible? Attempts at justification come comparatively late, but the fundamental grounding for moral responsibility lurks much deeper in our strikeback desires; we feel those desires so deeply that we are certain they *must* be justified.

The basic belief in moral responsibility is not a product of reason, but of visceral emotion. Obviously, contemporary philosophers who defend moral responsibility do not defend—and would deplore—attacks on scapegoats; to the contrary, the focus of their efforts is to make clearer and more precise exactly who is and is not morally responsible. But the deep belief in moral responsibility—among philosophers and the general public alike—is much stronger and deeper than the arguments for moral responsibility can begin to justify.

Moral responsibility was born on the wrong side of the tracks, in the harsh and indiscriminating strikeback reaction we feel when we suffer harm. That origin doesn't carry much philosophical weight. Its unsavory evolutionary history does not show that the basic emotional reaction underlying moral responsibility is bad. After all, our affection for our children probably has its origin in selfish aggressive genetic advantage; that self-serving history does not undercut the moral worth of our deep affection for our children. Primatologist Frans de Waal makes this point quite effectively:

Even if a diamond owes its beauty to millions of years of crushing pressure, we rarely think of this fact when admiring the gem. So why should we let the ruthlessness of natural selection distract from the wonders it has produced? Humans and other animals have been endowed with a capacity for genuine love, sympathy, and care—a fact that can and will one day be fully reconciled with the idea that genetic self-promotion drives the evolutionary process. (1996, 16–17)

Still, moral responsibility comes with a lot of baggage. Even if we grant that squalid origins do not establish that a value or emotion is wrong, it does not follow that origins are irrelevant to our judgments concerning values. Consider our aesthetic values. Many of us regard diamonds as beautiful, but when we learn that our aesthetic judgment was manipulated by diamond merchants through clever advertising campaigns (“a diamond is forever”) and an artificially contrived sense of something rare and wonderful (when in fact diamonds exist in quantities that would destroy their status as precious stones, were the mining not carefully limited), then diamonds lose some of their luster. And when we recognize that the deep and confident belief in moral responsibility is not based on solid rational arguments (because even the most committed supporters of moral responsibility acknowledge that their own belief in moral responsibility is stronger than any argument in favor of moral responsibility), then we may wish to look more skeptically at the primitive and unattractive visceral feelings that provide the ultimate support for tenacious belief in moral responsibility.

The point of noting the dark origins of moral responsibility is to combat the powerful presumption in favor of moral responsibility and encourage genuine scrutiny of that deep belief. Do we have good reasons for believing in the legitimacy of claiming and attributing moral responsibility, or do we instead start from the certainty that it is legitimate and so conclude that we must have reasons? When we look at moral responsibility from a thoroughly naturalistic perspective, do we really find it plausible? Or is moral responsibility an atavistic remainder of our evolutionary history: a remainder requiring miracles and deities that cannot be justified within our contemporary scientific system?

Systemic Belief in Moral Responsibility

The deep and almost unshakable belief in moral responsibility owes much of its strength to our retributive emotions and powerful strikeback desires. But there is another source for the constancy and certainty of our commitment to belief in moral responsibility: it is held securely in place by its central location in a larger system of belief, and challenging moral responsibility requires challenging that larger theoretical system.

C. A. Campbell (1957) attempts to square belief in moral responsibility with our scientific knowledge, but he insists that if that project were a failure, he would renounce his belief in the credibility of scientific research in order to preserve his belief in moral responsibility, as the latter is essential for our lives as practical ethical beings. P. F. Strawson (1962) claims that no argument is required to defend moral responsibility, because it is so central to our belief and value system that it cannot be denied without collapsing the foundation of our moral and emotional lives. Peter van Inwagen suggests that no matter what some philosophers may claim about their denial of moral responsibility, it is very unlikely that anyone genuinely rejects that deep belief. And van Inwagen affirms his unconditional commitment to moral responsibility: his strong doubts concerning compatibilism notwithstanding, if determinism were proved and the libertarian model were unsustainable, he would adopt compatibilism: "To deny the free-will thesis is to deny the existence of moral responsibility, which would be absurd. . . . It is conceivable that science will one day present us with compelling reasons for believing in determinism. Then, and only then, I think, should we become compatibilists" (1983, 223). Van Inwagen prefers to reject determinism and preserve a libertarian basis for moral responsibility, but if faced with problems he is willing to give up his libertarian commitments and embrace the despised compatibilism to save moral responsibility. As Quine made clear in "Two Dogmas of Empiricism," "Any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system" (1951, 40). And van Inwagen is willing to make whatever adjustments are necessary for the preservation of moral responsibility. But such a commitment to moral responsibility sounds more like a statement of faith than a reasoned conclusion: any possibility of rejecting moral responsibility is "absurd," and no doctrine is too implausible for acceptance if it saves moral responsibility.

Nick Trakakis and Daniel Cohen offer an interpretation of Wittgenstein according to which "belief in free will [and moral responsibility] begins to look more like a religious commitment than a theoretical or scientific belief" (2008, xvii). That is an interesting interpretation of belief in moral responsibility—as an article of faith—because it leads to the second reason why so many people (and perhaps especially philosophers) find the denial of moral responsibility too absurd to take seriously. Like religious doctrine, belief in moral responsibility is a central element of a complex and

comprehensive system of belief. That system frames the way we pose the questions and colors the way we see the arguments, and the assumptions of that system make it difficult to develop opposing arguments. The questions concerning moral responsibility are typically asked from within the framework of moral responsibility, so that rejection of moral responsibility soon leads to absurdity. Thus in order to challenge moral responsibility, one must challenge the vast moral responsibility legal and philosophical system; for those operating deep within that system, any challenge to the system is difficult to take seriously. From within that system (criminal justice is only one of its elements), the denial of moral responsibility does seem absurd; in such a rich system, it is always possible to save moral responsibility by making adjustments elsewhere in the system: for example, by significantly lowering the standards for justified blame until moral responsibility becomes a necessary truth, as Dennett proposes: “If no one *else* is responsible for your being in state A, you are” (2003, 281), or by ruling deeper inquiry into causal history out of order, or—as suggested by Stephen White—by starting from the assumption of justified moral responsibility and redefining what we mean by “could have done otherwise,” so that “an agent could have done other than he or she did just in case the ascriptions of responsibility and blame to that agent for the action in question is justified” (1991, 236).

Understanding the emotional and systemic sources of the almost universal belief in moral responsibility opens the way to careful consideration of the argumentative case in support of moral responsibility as well as the case that can be made against moral responsibility. There are, of course, arguments for moral responsibility; indeed, a multitude of conflicting arguments. The remainder of this book attempts to show that those arguments don’t work. But before examining the range of arguments in more detail, it should be noted that the great variety of arguments in support of moral responsibility is evidence of both the powerful emotional appeal of moral responsibility and the weakness of the philosophical arguments to support it. Many philosophers offer arguments to justify moral responsibility, each believing that the other arguments fail. If twelve jurors are convinced that the defendant is guilty, but each juror bases his or her guilty verdict on a single item of evidence that the other eleven jurors reject, that is not good grounds for conviction: for each element of evidence, eleven out of the twelve regard it as faulty. Ultimately, we would be convicting on evidence

that is almost unanimously rejected. The rich variety of arguments in favor of moral responsibility may remind us of the rich variety of arguments for the existence of God and may prompt the same response to that large collection: if there were really *a* good argument for God or for moral responsibility, would there be so *many*? In contrast to the multitude of arguments in support of moral responsibility, there is one basic argument against moral responsibility—though it is available in a number of different styles and colors. That argument is the subject of the next chapter.