The Moral Responsibility Delusion

The Moral Responsibility Delusion

By

Bruce N. Waller

Cambridge
Scholars
Publishing

The Moral Responsibility Delusion

By Bruce N. Waller

This book first published 2023

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-9016-X
ISBN (13): 978-1-5275-9016-8

Dedicated to my beloved wife, Mary Elizabeth Newell Waller.

“She openeth her mouth with wisdom; and in her tongue is the law of kindness.”

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I am deeply indebted to many philosophers with whom I have discussed moral responsibility, including Robert Kane, Richard Double, Gregg Caruso, Derk Pereboom, Farah Focquaert, Elizabeth Shaw, Michael Corrado, Dan Dennett, Saul Smilansky, Gerald Massey, Dane Gogoshin, Tom Clark, Peter Alces, Paul Davies, Michelle Ciurria, Ryan Lake, and John Lemos. Many of my colleagues at Youngstown State University greatly enriched my work, particularly Brendan Minogue, Thomas Shipka, Chris Bache, Alan Tomhave, Mark Vopat, Mustansir Mir, Deborah Mower, Linda "Tess" Tessier, Gabriel Palmer-Fernandez, Robyn Gaier, and Michael Jerryson. I was very fortunate to work with three fabulous department administrators: Joan Bevan, Mary Dillingham, and Linda Glover.

I have learned a great deal, and gained great enjoyment, from many years of lunches and coffees with very wise friends, including Fred Alexander, Lauren Schroeder, Luke Lucas, Jane Kestner, Jim Morrison, Tod Porter, Charles Singler, Ikram Khawaja, and Mark Womble. My friends Homer Warren and Linette Stratford have given me a rich new perspective on "producer consciousness."

My wonderful long-time friends from grad school at UNC-Chapel Hill – the notorious "McCauley Street Gang" – have been one of the steadfast joys of my life; their constant support has been invaluable.

My greatest debt is to my family: my son Russell and his wife Robyn, and their three delightful and brilliant children, Nathanael, Josephine, and Emil; my son Adam and his husband, Josh; and my remarkably patient, kind, and gentle wife, Mary. I dedicate this book to her.

INTRODUCTION

The traditional belief in moral responsibility is under stress. In addition to strong philosophical arguments against moral responsibility there are powerful challenges from research in psychology and neuropsychology. Philosophers have offered a remarkable variety of arguments to prop up moral responsibility, but the sheer number indicates that none have had great success in providing strong grounds for the embattled doctrine of just deserts. Despite the battering, moral responsibility remains firmly entrenched among both philosophers and the folk. The resilience of belief in moral responsibility suggests that there are hidden factors that are the actual basis for the stubborn belief in moral responsibility: factors that are either ignored by or invisible to the philosophical debate. By revealing and examining the deep and generally nonconscious forces that hold the moral responsibility system in place, *The Moral Responsibility Delusion* takes a unique approach to the question of moral responsibility. It exhibits the deep psychological forces that prop up belief in moral responsibility, examines the deceptive and typically nonconscious influence of those forces, and describes how the combination of misleading and misunderstood psychological influences sustains the implausible belief in moral responsibility.

Moral responsibility is often regarded as an indubitable intuition, a basic belief that is secure against all threats. *The Moral Responsibility Delusion* first examines the strength of belief in moral responsibility, and then critiques the wide variety of arguments that philosophers offer for the plausibility of moral responsibility: arguments from *causa sui* miracles, question-begging *reductio* arguments that treat the denial of moral responsibility as absurd, and arguments to justify blocking deeper inquiry into the causes of our behavior. The weakness of those arguments – in contrast to the confidence with which philosophers embrace belief in moral responsibility – raises questions concerning the *real* basis for the almost universal popular and philosophical commitment to moral responsibility. The answers to those questions are found in the examination of key psychological motives that typically operate beneath our consciousness. To the limited degree that we are conscious of those motives and their subtle influence, our interpretation of their nature and operations is deeply distorted. Each of those strong but nonconscious motives supports belief in moral responsibility, and when combined they support a destructive belief

system that is deeply entrenched but also deeply misunderstood. Examining the implausibility and misconceptions at the foundation of the moral responsibility system exposes the weakness of that foundation and opens a clearer view of the deep problems for belief in moral responsibility and just deserts. Eliminating the delusion of moral responsibility facilitates a better understanding of free will, better methods of shaping positive moral character and behavior, and a richer appreciation of vitally important moral sentiments.

CHAPTER ONE

DEDICATION TO MORAL RESPONSIBILITY

For many centuries, philosophers have devoted intense efforts to justifying belief in moral responsibility: from Aristotle to Epicurus, Lucretius to Augustine, Aquinas to Leibniz, Kant to van Inwagen, Kane to Korsgaard. The proposals have taken many forms, from the mundane to the exotic, but almost always there is an underlying sense that the champions of moral responsibility are seeking philosophical grounding for something everyone already knows to be true: that in normal circumstances we are morally responsible for our behavior. Indeed, some suspect that the difficulty in providing *proof* of moral responsibility is that something so obvious defies any attempt at deeper justification. Peter van Inwagen struggles to establish a solid basis for belief in moral responsibility; but it is purely an intellectual exercise, since moral responsibility is beyond any possible doubt: "I have listened to philosophers who deny the existence of moral responsibility. I cannot take them seriously" (1983, 207). For van Inwagen it is the indubitable truth of moral responsibility that is key to proving the existence of free will: "It is as adequate a defence of the free-will thesis as has ever been given for any philosophical position to say, 'Without free will, we should never be morally responsible for anything; and we are sometimes morally responsible'" (1983, 209).

Adina Roskies shares van Inwagen's fixed certainty concerning moral responsibility, categorizing moral responsibility as "the most vivid of our intuitions concerning human behavior":

> We ought to explore the possibility that freedom is a concept derivative on more robust intuitions about responsibility rather than vice versa. I favor a compatibilist view of responsibility, which has several advantages over incompatibilist views. It provides an account of responsibility that is not hostage to the outcome of the determinism/indeterminism debate, which, as I've argued, does not hold much promise for progress. Doing so preserves perhaps the most vivid of our intuitions about human behavior: that, given certain circumstances, we are responsible for our choices and actions (Roskies 2014, 121).

If the problem of free will troubles you, put your trust in the unshakable intuition of moral responsibility.

The Moral Responsibility Delusion

Two things are clear from the rich variety of philosophical efforts to find grounds for belief in moral responsibility: first, philosophers are sublimely confident that such grounds must be out there *somewhere*; and second, none of the many proposed answers have gained anything close to consensus support. So why is a doctrine that has little in the way of solid empirical or philosophical support so widely and fervently believed? It has powerful support from psychological forces that operate nonconsciously or deceptively. The real support for moral responsibility is based in deep delusion: in beliefs and motivations that are rarely – and sometimes never – available for conscious scrutiny. Among the most influential of those factors is our nonconscious *belief in a just world*. When considered consciously, it is obvious to almost everyone that the world is far from just. But belief in a just world typically operates below the level of conscious scrutiny, influencing in subtle ways our view of events and actions. A second misleading force is our powerful desire to "strike-back" at those who harm us: a desire that moral responsibility transforms into righteous retribution. Or so it seems. But we are deluded by the real motive. Rather than a desire to strike-back at transgressors, the real desire is to pass-the-pain-along to some convenient target, whether guilty or innocent. Another important deceptive influence that supports belief in moral responsibility is our systemic belief in the "folk metaphysics of agency" – our commonsense beliefs concerning our powers of deliberation and choice. Those beliefs are pleasing to our self-image but are in basic conflict with contemporary neuropsychological research. We believe that our deliberations and choices are transparent: that we are consciously aware of the factors that influence our thoughts and our decisions. But we are blissfully unaware of many of the important and even decisive influences on our deliberation and our choices. And we believe that it is always within our power to deliberate longer and harder, though there are many factors that influence – and often inhibit – both the length and depth of our deliberative capacity. Another source of distortion: We are most comfortable with teleological accounts of events, and quite adept at contriving such accounts. This deep tendency toward "promiscuous teleology" pushes us into false narratives that we find very appealing. When promiscuous teleology combines forces with the fundamental attribution error, the problem is exacerbated: we exaggerate our own cognitive and evaluative resources while ignoring the influence of

situational factors on our behavior. The roots and operations of the moral responsibility system are vastly different from what we imagine them to be, and the effects of the system are far from innocent. These misleading – even delusional – forces combine to make belief in moral responsibility seem indubitable. The nature of those forces and how they fit together in support of the moral responsibility system is the focus of this book.

The Strength of the Moral Responsibility System

Afflicting philosophers and folk alike, the moral responsibility delusion is powerful. Its hold on the folk is at least as strong as its grip on philosophers. Attempts to challenge moral responsibility are likely to be met with a decisive rejection: "Everyone knows the difference between right and wrong, and everyone has the ability to choose the right path." Those who make that claim may not characterize this remarkable ability of absolute choice as an appeal to a godlike first cause, but that is precisely what it is: nothing in your past or your environment or your circumstances has any effect whatsoever; your act of choosing between right and wrong has no causal antecedents or conditions.

This belief in moral responsibility is so strong that most people have a difficult time even imagining its rejection. In a fascinating experiment by Shaun Nichols and Joshua Knobe (2007), experimental subjects read an account of a completely deterministic universe (Universe A). One group was then asked an abstract question concerning moral responsibility and determinism: "In Universe A, is it possible for a person to be fully morally responsible for their actions?" Fewer than 5% of the subjects believed persons could be fully morally responsible in that deterministic universe. The second group were told a disturbing story of a man who kills his wife and three children to marry his attractive secretary. In the second case – operating with precisely the same deterministic Universe A – 72% of the subjects concluded that the murderer was fully morally responsible for his actions. Faced with a concrete case arousing strong emotions – a cruel murderer – careful deliberation and deeper inquiry are blocked (Sherman, Beike, and Ryalls 1999; Hafer and Bègue 2005, 137) and the moral responsibility of the murderer seems obvious.

It has long been assumed that the conception of free will among ordinary folks – those not corrupted by philosophy – is some form of *libertarian* free will that is *incompatible* with determinism. Precisely what that libertarian free will is might be open to question (perhaps the folk, not thinking so obsessively about the nature of free will, are not themselves clear on the details); but the folk conception of free will has generally been

assumed (by philosophers) to involve some form of special creative *choice* among open alternatives, and the resulting choice is entirely and exclusively under the control of the chooser and is not determined by any force or cause other than the chooser's own special choice. Recent studies by philosophers and psychologists have called into question the assumption that ordinary folk favor such a libertarian version of free will: those studies offer evidence that the ordinary understanding of free will is *compatibilist*. Based on those studies, some conclude that folk free will (and the moral responsibility it is assumed to support) is *compatible* with belief in determinism: determinism poses no threat to the popular understanding of free will.

In a classic experiment on this subject (Nahmias, Morris, Nadelhoffer, and Turner 2005) researchers set up a determinist scenario (for example, a case in which a supercomputer combining knowledge of all the laws of nature together with complete knowledge of the current state of everything in the world is able to predict every detail of the future with 100% accuracy), and then asked subjects if a person performing an act – such as robbing a bank – in that deterministic world would have acted of his or her own free will and be morally responsible for the act. Better than three quarters of the participants concluded that the actor in the deterministic world acted with free will and is morally responsible for the act.

While such research may give some support for the view that folk free will is compatibilist, there are strong challenges to that conclusion. A more likely scenario – supported by subsequent research – is that under the folk conception of moral responsibility there can be no moral responsibility in a deterministic world. But a *concrete case* such as robbing a bank – or as in the Nichols and Knobe (2007) experiment, the murder of an innocent woman and her three children – triggers powerful emotions that overwhelm careful deliberation and blow away any hypothetical implications of a deterministic world.

Hypothetical considerations of deterministic scenarios are mother's milk to philosophers; they are strange and alien to most experimental subjects. When subjects try to picture a deterministic world, they are likely to wind up with a picture of a fatalistic world in which they are totally helpless, and all the events in that world happen independently of their own thoughts and efforts: a world, in Alfred Mele's phrase (1995, 182-185), that totally *bypasses* them and makes them helpless observers of what happens (or even worse, makes them helpless puppets of external forces that control them). Another possibility is that many subjects asked to imagine themselves in a *deterministic* universe have a difficult time practicing a "willing suspension of disbelief" in the denial of their cherished libertarian free will, and even when they try to imagine determinism they cling to exceptions in the

proposed deterministic model. For example, they may adopt a Cartesian exemption from total determinism: perhaps the physical world is completely determined; but there is still room for *mental* freedom in the form of special deliberative choices (which they believe could not *possibly*, not even in any *imaginable* world, be determined). As Robert M. Sapolsky notes: "One of the deepest lines drawn by believers in mitigated free will . . . is the belief that one assigns aptitude and impulse to biology and effort and resisting impulse to free will" (2017, 596).

A study of the impact of "intrusive metaphysics" on the imaginative powers of research subjects – subjects attempting to picture a deterministic world (Nadelhoffer, Rose, Buckwalter, and Nichols 2020) – found that "intuitive views about the indeterministic nature of human agency influence how people understand deterministic cases . . . when they are making free will judgments. Owing to this indeterministic intrusion, most people have a difficult time adequately tracking the determinism built into two of the central scenarios that have been used to support natural compatibilism" (2020, 16). The researchers concluded there are *flaws* in the experiments (the experiments purportedly showing the folk believe that free will is compatible with determinism): those experiments are invalidated by the *intrusion* of free will beliefs that are inconsistent with determinism:

> We believe that the main reason that intrusion effects are prevalent is that indeterminism about human agency is the default folk view. That so many participants find deterministic scenarios to be impossible highlights the intuitive appeal of an indeterministic metaphysics. It also presumably partly explains why some people appear to have such a hard time tracking determinism (Nadelhoffer, Rose, Buckwalter, and Nichols 2020, 17).

That research built on earlier work (Rose, Buckwalter, and Nichols 2017), in which a series of six experiments provided "evidence that participants *fill in* the [deterministic neuro-prediction] scenarios in ways that undermine the inference that the ordinary notion of free will is compatible with the idea of perfect neural prediction." The experimenters concluded that:

> We found that participants overwhelmingly attribute free will in cases of [deterministic] perfect neuro-prediction. However, we also found that participants' intuitive metaphysics of free will intrudes into their representation of these perfect neuro-prediction scenarios . . . . The presence of intruding and importing effects suggests that people are imposing an indeterminist notion of free will onto the situation, despite the

> fact that the situation is explicitly described in terms of perfect predictability. Thus, we doubt that people are broadly comfortable with the idea of perfect neuro-prediction or that it is fully compatible with commonsense notions of free will (Rose, Buckwalter, and Nichols 2017, 499).

Not only is an incompatibilist free will the ordinary belief, but it is so strong that it "intrudes" into efforts to fully imagine a determinist scenario (see also Clark, Winegard, and Baumeister 2019). The "folk" are not merely strong believers in a special free will that conflicts with determinism; that belief is so resilient that it blocks efforts to even consider an alternative perspective that rejects such free will.

One other measure of the strength of belief in moral responsibility – and the special power of self-making choice that supports it – can be found in religious belief. Scripture is clear: humans have no such choice; omnipotent God holds all the power of choice, and God fashions humans according to His inscrutable wishes. Some are chosen for grace while the rest spend eternity in hell. God has *all* the power, and – as Paul makes clear – humans have no say or choice in the outcome.

> Nay but, O man, who art thou that repliest against God? Shall the thing formed say to him that formed it, Why hast thou made me thus? Hath not the potter power over the clay, of the same lump to make one vessel unto honour, and another unto dishonour? (Romans 10:20-21)

Why is one made vile, and the other given grace? It is totally God's choice, and there is no human free will:"So then it is not of him that willeth, nor of him that runneth, but of God that sheweth mercy" (Romans 10:16). This deeply offended Martin Luther's sense of justice and fairness, but his faith pulled him through:

> This is the highest degree of faith – to believe that He is merciful, who saves so few and damns so many; to believe Him just, who according to His own will makes us necessarily damnable . . . . If, therefore, I could by any means comprehend how that same God can be merciful and just who carries the appearance of so much wrath and iniquity, there would be no need of faith. But now, since that cannot be comprehended, there is room for exercising faith . . . (1525/1823, section 24).

The idea of burning in hell is disturbing, but the notion that we have absolutely no control over our destiny makes it even worse. Even the fundamentalists – perhaps especially the fundamentalists – found that

unacceptable, whatever Scripture might say. Scripture is sacred, literally true in every detail; but we *know* we have special powers of choice, and we know that God would not condemn us unjustly, and we know that blame and punishment could not be just in the absence of special powers of free will. Orthodoxy and Scripture may condemn the notion that we have godlike free will, but we are made in the image of God; and if God has miraculous free will, why shouldn't we?

Butcher, baker, or philosopher, belief in moral responsibility is so powerful it seems indubitable. It controls our social practices, is built into our justice system, dramatized in song and story and Western movie and crime novel, operates deep in our religious beliefs (even when it must defeat orthodoxy), is part of the American myth of self-made men and rugged individuals and the belief that everyone has genuine opportunity, and it is deep in our ethics (everyone knows the difference between right and wrong and has the special power to freely choose). It is not surprising that the moral responsibility system seems above challenge.

Folk Intuitions

In defending the importance of philosophy experiments to determine the view of the "folk" (those untutored in philosophy) on the question of moral responsibility, Adina Roskies and Shaun Nichols claim that: "Consistency with the deliverances of folk intuitions is a sign that the philosopher is making contact with his object of interest" (2008, 371). It is clearly important to make contact with one's object of interest: if the folk seem to be talking about something totally foreign to any questions related to moral responsibility as I understand it, that is a reason to think that my questions are misplaced. But if "consistency with the deliverances of folk intuitions" means something stronger – that the "deliverances of folk intuitions" have a special status as fundamental truths – then that claim has little plausibility. It may well be that the folk – and even some philosophers! – are deluded by the powerful influences of tradition, culture, and nonconscious beliefs and attitudes, and that their beliefs concerning moral responsibility are a product of that delusion. There is a precedent for such a possibility. In the early 19th Century, belief in divine design of species was deeply intuitive, not only for the folk but also for the most respected biological scientists. Had Darwin confined his theorizing to views consistent with the intuition of divine design, he could not have proposed the theory of natural selection. Belief in divine design was deeply intuitive, but we can study it and understand why it became entrenched and why – even if once valuable – it had become an impediment to better understanding. Moral responsibility is a similar

impediment, and this book is an attempt to understand the delusions that hold it in place.

If we step outside the constraints of the moral responsibility system and examine it closely, its implausibility is striking. Within the system, it cannot be doubted; stripped of the system, it cannot be believed. Moral responsibility is atavistic, a relic of a bygone era. Moral responsibility fits comfortably with miraculous godlike powers that make just deserts seem fair and plausible. One might have expected that the demise of gods and miracles and mysteries would have destroyed philosophical support for moral responsibility. But belief in moral responsibility was too strong, for reasons the remainder of this book will explore. Philosophers – always resourceful when defending lost causes – embraced the defense of moral responsibility with a remarkable variety of arguments. We might have hoped that philosophers would acknowledge that moral responsibility has lost all plausibility and would focus on finding a better alternative. Instead almost all philosopher devoted themselves to finding some way to save moral responsibility. The failure of those efforts – described in Chapters Two through Five – is an indication of the depth of the moral responsibility delusion. The sources of that delusion are examined in Chapters Six through Ten. The possibility of eliminating the moral responsibility delusion – and the benefits that would follow, from a better understanding of free will to a more *nearly* just criminal justice system – are described in the final three chapters.

CHAPTER TWO

CAUSA SUI

There is a simple, obvious, ancient, and deeply appealing basis for moral responsibility: we are morally responsible for our characters and our behavior because we *make ourselves* by our own choices, and those choices have no causal antecedents and no further explanation beyond our own *first cause* free choices. Ultimately, we are *causa sui*: we are responsible for ourselves because we made ourselves what we are. That's what it feels like, and the belief is so strong that most people cannot seriously question or challenge it. Even philosophers who insist they reject *causa sui* powers and spurn supernatural forces often appeal to special human powers that seem more appropriate for deities than mortal animals. Punishment is justly deserved because we are *morally responsible* for the behavior punished and the character which generates that behavior; and ultimately we are morally responsible because the crucial character traits are self-made.

Causa sui choices remain the gold standard for establishing moral responsibility, but they encounter an insurmountable obstacle. With miracles you can "explain" anything, but such convenient "explanations" are vacuous. Any account that includes the line "and then a miracle occurs" is unlikely to gain much support among contemporary philosophers. But contemporary philosophers are not immune to the charms of godlike powers of self-creation. Roderick Chisholm describes quite clearly the miraculous nature of such libertarian choice:

> If we are responsible, and if what I have been trying to say is true, then we have a prerogative which some would attribute only to God: each of us, when we really act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing and no one, except we ourselves, causes us to cause those events to happen (1982, 32).

Unfortunately for *causa sui* aspirations, Nietzsche made the problems – even the incoherence – of *causa sui* painfully clear:

> The *causa sui* is the best self-contradiction that has been conceived so far, it is a sort of rape and perversion of logic; but the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. . . . The desire to bear the entire and ultimate responsibility for one's actions . . . involves nothing less than to be precisely this *causa sui* and . . . to pull oneself up into existence by the hair, out of the swamps of nothingness (Nietzsche 1886/1966, section 21).

Nietzsche makes plain the basic *causa sui* problem: before the self is made, *who* is doing the making? How does one "pull oneself into existence" before one exists?

Fair Starts

Daniel Dennett would scoff at the notion that one could "pull oneself up into existence by the hair, out of the swamps of nothingness," yet Dennett is drawn to the idea of self-making as grounds for moral responsibility:

> I take responsibility for any thing I make and then inflict upon the general public; if my soup causes food poisoning, or my automobile causes air pollution, or my robot runs amok and kills someone, I, the manufacturer, am to blame. And although I may manage to get my suppliers and subcontractors to share the liability somewhat, I am held responsible for releasing the product to the public with whatever flaws it has. Common wisdom has it that much the same rationale grounds personal responsibility; I have created and unleashed an agent who is myself; if its acts produce harm, the manufacturer is held responsible. I think this common wisdom is indeed wisdom . . . (1984, 85).

This seems like common wisdom when we are operating within the moral responsibility system, but closer inspection soon uncovers fatal problems. Each "manufacturer" of a human agent starts with radically different raw materials, workshops that range from impoverished to luxurious, and skill levels that vary from near incompetence to superbly trained. With such dramatically different starting resources, moral responsibility for the end product appears to be based on myopic refusal to look deeper rather than any sort of wisdom.

Dennett, however, is aware of the problem of starting points, and he struggles to deal with it. If self-makers start with workshops, skills, and raw materials that are dramatically different, then holding them morally

responsible for the result is unfair. But if the differences are trivial, and – all things considered – the initial advantages and disadvantages even out, then the fair start results in morally responsible finished products that can be judged in terms of their just deserts:

> Suppose – what certainly seems to be true – that people are born with noticeably different cognitive endowments and propensities to develop character traits. . . . Is this 'hideously unfair' . . . or is this bound to lead to something hideously unfair? Not necessarily.
>
> Imagine a footrace in which the starting line was staggered: those with birthdays in January start a yard ahead of those born in February, and eleven yards ahead of those born in December. Surely no one can help being born in one month rather than another. Isn't this markedly unfair? Yes, if the race is a hundred yard dash. No, if it's a marathon. In a marathon such a relatively small initial advantage would count for nothing, since one can reliably expect other fortuitous breaks to have even greater effects . . . . Is it fair enough not to be worth worrying about? Of course. After all, luck averages out in the long run (1984, 95).

The child born to poverty and environmental hazards and substandard schools – and perhaps neglect or even abuse – may start out behind, but those disadvantages will be balanced out by compensating good fortune during childhood and adolescence. Only a belief in karma or fairy godmothers could support such a fantasy. Dennett knows that it is *not* the case that "luck averages out in the long run." Instead, "early-life advantages and disadvantages persist into late life, despite the many intervening contingencies and random life events. Indeed, rather than attenuating over time, economic effects of early advantages are often magnified over the life course" (Crystal, Shea, and Reyes 2017, 911). The cumulative advantages of being born into affluence (better health care, superior schools, advantageous connections, and legacy admissions to the most prestigious universities) are as obvious as the cumulative disadvantages of childhood poverty (from lead poisoning to malnutrition to inferior schools) (Zimmer and House 2003; Crystal and Shea 2002; DiPrete and Eirich 2006; Wolff and de-Shalit 2007; Raine 2013; Giudici and Pallas 2014). As Wolff and de-Shalit demonstrate, many of the disadvantages *cluster* together in ways that exacerbate their effects (2007, 119-128). The deep belief in just deserts blinds us to the inequity and injustice that destroy the basis for moral responsibility.

Like Dennett, George Sher is keenly aware that making moral responsibility plausible requires that people have at least roughly equal opportunities for success. And like Dennett, he knows full well that there is great variation in starting resources. For Dennett, luck averages out; for Sher, those with weak resources in one area can level the playing field through stronger resources in other areas, so there is actually a fair distribution of talents and abilities:

> Even if M is initially stronger or more intelligent than N, this difference will only entail that M does not deserve what he has achieved relative to N if the difference between them has made it impossible for N to achieve as much as M. However, differences in strength, intelligence, and other native gifts are rarely so pronounced as to have this effect. The far more common effect of such differences is merely to make it more *difficult* for the less talented person to reach a given level of attainment. He must work harder, husband his resources more carefully, plan more shrewdly, and so on (1987, 31-32).

We love stories of the less talented achieving victory through superior fortitude or clever strategy: the persevering plodding tortoise outraces the swift but lethargic hare. But we all know the sad truth we hate to acknowledge: the lottery of life does not distribute its prizes in an even-handed manner. To the contrary, strengths and weaknesses mesh together. As Wolff and De-Shalit (2007, 119-128) show through their research, many of our initial disadvantages "cluster" together, becoming powerful forces that often have "corrosive" effects on future prospects. The initially brighter student receives more encouragement and more personal attention, is positively reinforced and gains confidence from her successes, is enrolled in special classes and participates in special projects. The deeply discouraged student simply *cannot* work harder, while the positively reinforced student eagerly does so. That everyone can always work harder and try harder is a charming but destructive myth. Like reason, the capacity to "try harder" is godlike. Dennett and Sher recognize that if self-making is to be plausible as grounds for moral responsibility, then there must be fair starts. But their efforts to overcome that problem place in bold relief the profoundly divergent starting points that undercut self-making as grounds for moral responsibility.

Robert Kane's First Causes

We are morally responsible because we make ourselves, we are *causa sui*. That defense of moral responsibility has deep appeal, but even deeper problems. But its charms are such that champions of moral responsibility do not abandon self-making without a struggle. Rather than ultimate *ab initio* self-making, some suggest that sufficient self-making can occur at some crucial point: some point at which the self-making choice has no causal antecedents. Easy enough, if we have godlike first cause creative powers; not so easy if we are mortals shaped by our genetics and environments. Robert Kane makes a fascinating effort to find a first cause starting point for critical self-making decisions while scrupulously avoiding any appeal to the miraculous and abjuring "mysterious accounts of human agency that are empty of explanatory content" (1985, 12). Kane seeks special acts of free will that have no explanation or cause prior to the act itself and the choice of the agent:

> To have free will therefore is to be the ultimate designer of one's own purposes or ends or goals. And if we are to be the ultimate designers of our own purposes or ends, there must be *some* actions in our life histories that are will-setting, plural voluntary [there are several genuine open options one can choose] *and* undetermined by someone or something else (2007, 22).

Kane's elaborate system for generating freely willed *first* causes – without miracles – focuses on special "self-forming actions" (2007, 14): character-forming moral choices that occur when there is a conflict between what one desires and one's sense of what is morally right. In such pivotal situations the person experiences turmoil and engages in intense deliberation concerning the choice. The intense deliberation stimulates the brain and results in chaotic events that amplify the indeterminacy of a subatomic particle, and that indeterminacy becomes part of the decision-making process. Because indeterminism is a crucial element of the resulting moment of choice, that specific moment becomes the final *stopping* point for any explanation: no inquiry can go deeper, no earlier causes can be traced beyond the indeterminacy limit, and the choosing individual is a genuine first cause. Whichever choice the indeterminacy marker falls on, the choice is one that had strong appeal to the chooser – is actually *chosen* by the chooser – and this special choice belongs *ultimately* to the chooser: indeterminacy eliminates deeper causes.

Kane's ingenious effort to save moral responsibility is an important lesson in the challenges confronting any attempt at a *causa sui* basis for

moral responsibility. There may be legitimate doubts concerning the success of Kane's project: there are many moving parts in Kane's model of special self-forming choices, and some of the parts may be less plausible than others; and even if Kane manages to make a strong case for the process, there is still the problem of how this can be *my* choice if it resulted from *chance*: that doesn't sound like my choice at all. But whatever the plausibility of Kane's model for supporting moral responsibility, it is an honest effort to face the full *challenge* of making sense of ultimate *causa sui* responsibility.

Adina Roskies' Self-Authorship

Another attempt at using special self-making choices to ground moral responsibility has been made by Adina Roskies. Like Kane, she rejects any miraculous appeals; and like Kane, she clearly recognizes that to make moral responsibility plausible there must be a point at which no deeper inquiry is possible because "the buck stops here" (Roskies 2012, 333). Roskies acknowledges that we start with varying abilities and desires, but she argues that through reform and remodeling one can "make oneself" to such a degree that the resulting person is genuinely morally responsible for who she or he is. As we make our deliberative *choices* we gradually reshape ourselves in accordance with those choices.

Roskies' version of self-making (she calls it self-authorship) emphasizes a long process of self-making that is heavily dependent on careful deliberation:

> Through deliberately thinking and acting in strategic ways we can exert control, modulate and intervene in our future states, both physical and mental. It is this that allows us to shape ourselves in ways that make itthe case that we are in some very real sense responsible for who we are (2012, 331).

Roskies recognizes Nietzsche's basic problem of self-making: how can we make ourselves out of nothing? Who is doing the construction, and by what plan? Roskies insists that "intermediate" self-making – with the aid of strong deliberation – avoids those problems.

> This need not be the self-making from the ground (or the swamp) up: It is not necessary to create oneself out of whole cloth. Continual shaping of the given over time establishes sufficient control of the agent to suffice as a grounding kind of self-causation (2012, 338).

But that places a sharp focus on the nature of such self-reforming choices. For Roskies, what makes such choices the appropriate stopping point and grounds for moral responsibility is that they are made by our own special conscious cognitive powers:

> I am not denying that there will always be some kind of non-reason, physical explanation for the presence of a reason, but what I am suggesting is that that is not the place to look. Certainly, our reasons may be elements of us that are in some sense given. But they are not unchangeable. The agent, in appreciating the content of a reason, can intervene in the way in which that reason operates in his deliberation, in loops of agential control. If he or she did not, and by not doing so the agent accepted and endorsed that reason and its role in his deliberation, then . . . that is the appropriate place to stop. Agents control themselves via conscious access to mental content. We therefore have a convenient way of locating the source of the control: at the agentive, aware level (2012, 334-335).

For Roskies, that is the appropriate place to stop because it fits "the folk metaphysics of agency" (2012, 35). But the folk metaphysics of agency is just the old belief in moral responsibility combined with an exaggerated and distorted view of human cognition. There is no more reason to trust the folk metaphysics of agency than there is to support the folk metaphysics of creationism. Our intuitions about our cognitive functioning seem especially trustworthy because they seem so intimate: we exercise control "via conscious access to mental content." Sometimes we do manage to exercise some effective degree of conscious control, but such occasions are rarer than we suppose, and the causal factors are more complex than we imagine: the factors of which we are consciously aware are only a small part of our deliberations.

The Folk Metaphysics of Agency

The folk metaphysical belief that "Agents control themselves via conscious access to mental content" is charming but false. The "agentive aware level" is blind to many important factors. Psychological research (Wilson 2002; Kahneman 2011; Panksepp and Biven 2012; Dehaene 2014; Davies 2019) has made that abundantly clear, but the prophet Jeremiah did not require those fascinating studies to recognize that our agentive deliberative decisions are often blind to decisive causal factors and that we often deceive ourselves concerning our own motives and reasons: "The heart is deceitful

above all things, and desperately wicked: who can know it?" (Jeremiah 17:9 KJV) What Jeremiah knew millennia ago contemporary psychologists have confirmed and elaborated in important ways. Rather than agentive unbiased fully conscious deliberation determining our choices, much of our deliberation is rationalization to justify what we wish to do for nonconscious emotional motives. Jonathan Haidt describes the actual process with a striking metaphor:

> Automatic processes run the human mind, just as they have been running animal minds for 500 million years, so they're very good at what they do, like software that has been improved through thousands of product cycles. When human beings evolved the capacity for language and reasoning at some point in the last million years, the brain did not rewire itself to hand over the reins to a new and inexperienced charioteer. Rather, the rider (language-based reasoning) evolved because it did something useful for the elephant. The rider … acts as the spokesman for the elephant, even though it doesn't necessarily know what the elephant is really thinking. The rider is skilled at fabricating post hoc explanations for whatever the elephant has just done, and it is good at finding reasons to justify what the elephant wants to do next. Once human beings developed language and began to use it to gossip about each other, it became extremely valuable for elephants to carry on their backs a full-time public relations firm (2012, 45–46).

Haidt and Jeremiah make clear the implausibility of an "agentive, aware level" at which "agents control themselves via conscious access to mental content." There are many factors that influence and often distort the reasoned deliberation we imagine to be unbiased and transparent and totally under our special conscious control. First, the capacity for engaging and sustaining our "System 2" (Kahneman 2011) powers of careful deliberation varies widely. It is a popular misconception that almost everyone can always think longer and harder and deeper, and it is a misconception that is especially popular among philosophers. The eminent philosopher Charles Taylor celebrated our ubiquitous powers of radical self-evaluation and "self-resolution":

> Because this self-resolution [deep self-evaluation] is something we do when we do it, we can be called responsible for ourselves; and because it is within limits always up to us to do it, even when we don't – indeed, the nature of our deepest evaluations constantly raises the question of whether we have them right –

> we can be called responsible in another sense for ourselves whether we undertake this radical evaluation or not (1976, 299).

Taylor's view that "it is within limits always up to us to do it [deep self-evaluation], even when we don't" is philosophical fantasy. There are large differences in the capacity for System 2 deliberation: differences in what researchers label "need for cognition" (Cohen, Stotland, and Wolfe 1955). "Chronic cognizers" engage more frequently in deliberative efforts and sustain them longer, while "cognitive misers" find System 2 deliberation more aversive and abandon it quickly (Cacioppo and Petty 1982). Not only are there large differences in need for cognition but also important variations in *confidence* in one's ability to reason effectively. Albert Bandura describes differences in *cognitive self-efficacy* (Bandura 1977): the degree of confidence in one's ability to successfully solve problems and discover better paths by careful deliberation. After repeated failures in vigorous cognitive efforts (such as early childhood attempts at difficult cognitive challenges that were beyond the child's abilities) cognitive self-efficacy is severely weakened. The frustrating student who "won't even try" to solve a challenging logic problem may be incapable of making such an effort. The "willful refusal" to exert effort is obvious; the unfortunate history that shaped such cognitive helplessness is invisible.

When Jonathan Haidt gave research subjects cases in which they already favored a specific ethical conclusion, but in which all the grounds in support of their favored view were conclusively blocked, the subjects would reluctantly acknowledge the failure of their given reasons for their favored view but still insist that their view was correct and that there *must* be good reasons in support of it. As Haidt describes the uncomfortable "reasoning" of the subjects:

> These subjects were reasoning. They were working quite hard at reasoning. But it was not reasoning in search of truth; it was reasoning in support of their emotional reactions (2012, 25).

If we do engage in "self-resolution" it is more likely to be a rationalization of our values than a transparent objective deliberative evaluation.

Roskies' confident assertion that "Agents control themselves via conscious access to mental content" is a common but false view of our cognitive operations. Michael Gazzaniga's view is jarring to our "common sense" beliefs but more accurate: "Our conscious awareness is the mere tip of the iceberg of nonconscious processing" (2011, 68). Researchers have confirmed the powerful force of nonconscious cognitive operations, but such operations are also demonstrated by common experience. We struggle

to remember a name and our conscious thought dismisses the query, only to have the nonconscious cognitive gears spit out the name when we "are no longer thinking about it." Psychological research reveals the profound but almost entirely nonconscious influence of *belief in a just world* (Lerner 1980). Stereotyping (Kunda and Spencer 2003; Nosek et al. 2007) is especially destructive because it typically operates without our conscious awareness (and often despite our conscious denial). Unconscious confirmation bias can confound even careful researchers. Seemingly trivial *situational* factors – such as the encouragement of an authority figure, who influences two-thirds of subjects to inflict what they thought might be a fatal series of high voltage shocks (Milgram 1963) – profoundly influence our thought and behavior, causing us to perform acts that we *consciously* insist we would not do under any circumstances.

Roskies is concerned that we not push inquiries into "irrelevant" considerations of neuropsychology. In fact, such studies are of great relevance for understanding our cognitive processes, which involve a wide range of causes of which we are never conscious. As Jaak Panksepp notes:

> Although arousals of the primary-process emotional networks of mammalian brains are intensely experienced by humans and other animals, it is especially important to recognize that the secondary processes of the BrainMind, the basic forms of learning, memory and habit formation, are among the most unconscious "mental" processes of them all. . . . Our primal emotional needs and bodily motivations shape who we become before we know – before we become "aware of" – what is happening cognitively, often yielding end results without our "personal" consent (Panksepp and Biven 2012, 448).

One neuropsychological researcher commented that our kidneys produce urine and our brains produce thought, and introspection yields no more knowledge of how the latter functions than the former. As Timothy D. Wilson notes: "The causal role of conscious thought has been vastly overrated; instead, it is often a post-hoc explanation of responses that emanated from the adaptive unconscious" (Wilson 2002, 107). And as Jonathan Miller elaborates: "Human beings owe a surprisingly large proportion of their cognitive and behavioral capacities to the existence of an 'automatic self' of which they have no conscious knowledge and over which they have little voluntary control" (1995; cited in Wilson 2002, 73). This is dead contrary to the "folk metaphysics of agency," which represents our cognitive processes as perfectly transparent.