1. Introduction

Trial by ordeal seemed to me, as I learned about it in school, ridiculously unfair. How could it have endured as an institution in Europe for hundreds of years? The central idea was simple: with God's intervention, innocence would plainly reveal itself, as the accused thief sank to the bottom of the pond, or the accused adulterer remained unburned by the red hot poker placed in his hand. Only the guilty would drown or burn. (For witches, the ordeal was less "forgiving": if the accused witch drowned she was presumed innocent; if she bobbed to the surface, she was guilty, whereupon she was hauled off to a waiting fire.) With time on our hands, my friend and I concocted a plan. She would falsely accuse me of stealing her purse, and then I would lay my hand on the stove and see whether it burned. We fully expected it would burn, and it did. So if the test was that obvious, how could people have trusted to trial by ordeal as a system of justice?

From the medieval clerics, the answer would have been that our test was frivolous, and that God would not deign to intervene with a miracle for the benefit of kids fooling around. That answer seemed to us

a bit cooked up. What is the evidence God *ever* intervened on behalf of the wrongly accused? A further difficulty concerned nonbelievers, such as those not yet reached by missionaries, or . . . maybe *me*? Still, this answer alerted us to the matter of metaphysical (or as we said then, "otherworldly") beliefs in moral practices, along with the realization that what seemed to us obvious about fairness in determining guilt might not be obvious after all.

My history teacher tried to put the medieval practice in context, aiming to soften slightly our sense of superiority over our medieval ancestors: in trial by ordeal, the guilty were more likely to confess, since they believed God would not intervene on their behalf, whereas the innocent, convinced that God would help out, were prepared to go to trial. So the system might work pretty well for getting confessions from the guilty, even if it did poorly for protecting the innocent. This answer alerted us to the presence of pragmatics in moral practices, which struck us as a little less lofty than we had been led to expect. How hideously *unfair* if you were innocent and did go to trial. I could visualize myself, bound by ropes, drowning in a river after being accused of witchcraft by my piano teacher.[1]

So what is it to be fair? How do we *know* what to count as fair? Why do we regard trial by ordeal as *wrong*? Thus opens the door into the vast tangled forest of questions about right and wrong, good and evil, virtues and vices. For most of my adult life as a philosopher, I shied away from plunging unreservedly into these sorts of questions about morality. This was largely because I could not see a systematic way through that tangled forest, and because a lot of contemporary moral philosophy, though venerated in academic halls, was completely untethered to the "hard and fast"; that is, it had no strong connection to evolution or to the brain, and hence was in peril of floating on a sea of mere, albeit confident, opinion. And no doubt the medieval clerics were every bit as confident.

It did seem that likely Aristotle, Hume, and Darwin were right: we are social by nature. But what does that actually mean in terms of our

brains and our *genes*? To make progress beyond the broad hunches about our nature, we need something solid to attach the claim to. Without relevant, real data from evolutionary biology, neuroscience, and genetics, I could not see how to tether ideas about "our nature" to the hard and fast.

Despite being flummoxed, I began to appreciate that recent developments in the biological sciences allow us to see through the tangle, to begin to discern pathways revealed by new data. The phenomenon of moral values, hitherto so puzzling, is now less so. Not entirely clear, just less puzzling. By drawing on converging new data from neuroscience, evolutionary biology, experimental psychology, and genetics, and given a philosophical framework consilient with those data, we can now meaningfully approach the question of where values come from.

The wealth of data can easily swamp us, but the main story line can be set out in a fairly straightforward way. My aim here is to explain what is probably true about our social nature, and what that involves in terms of the neural platform for moral behavior. As will become plain, the platform is *only* the platform; it is not the whole story of human moral values. Social practices, and culture more generally, are not my focus here, although they are, of course, hugely important in the values people live by. Additionally, particular moral dilemmas, such as when a war is a just war, or whether inheritance taxes are fair, are not the focus here.

Although remarks of a general sort concerning our nature often fall on receptive ears, those same ears may become rather deaf when the details of brain circuitry begin to be discussed. When we speak of the possibility of linking large-scale questions about our mind with developments in the neurosciences, there are those who are wont to wag their fingers and warn us about the perils of *scientism*. That means, so far as I can tell, the offense of taking science into places where allegedly it has no business, of being in the grip of the grand delusion that science can explain everything, do everything. Scientism, as I have been duly wagged, is overreaching.

The complaint that a scientific approach to understanding morality commits the sin of scientism does really exaggerate what science is up to, since the scientific enterprise does not aim to displace the arts or the humanities. Shakespeare and Mozart and Caravaggio are not in competition with protein kinases and micro RNA. On the other hand, it is true that philosophical claims about the nature of things, such as moral intuition, are vulnerable. Here, philosophy and science are working the same ground, and evidence should trump armchair reflection. In the present case, the claim is not that science will wade in and tell us for every dilemma what *is* right or wrong. Rather, the point is that a deeper understanding of what it is that makes humans and other animals social, and what it is that disposes us to care about others, may lead to greater understanding of how to cope with social problems. That cannot be a bad thing. As the Scottish philosopher Adam Smith (1723–90) observed, "science is the great antidote to the poison of enthusiasm and superstition." By *enthusiasm* here, he meant *ideological fervor*, and undoubtedly his observation applies especially to the moral domain . Realistically, one must acknowledge in any case that science is not on the brink of explaining everything about the brain or evolution or genetics. We know more now than we did ten years ago; ten years hence we will know even more. But there will always be further questions looming on the horizon.

The scolding may be sharpened, however, warning of the logical absurdity of drawing on the biological sciences to understand the platform for morality. Here the accusation is that such an aim rests on the dunce's error of going from an *is* to an *ought*, from *facts* to *values*. Morality, it will be sternly sermonized, tells what we *ought* to do; biology can only tell what *is* the case.[2] With some impatience, we may be reproached for failing to heed the admonition of another eighteenth-century Scottish philosopher, David Hume (1711–76), that you cannot derive an *ought* statement from statements about what *is*. Hence my project, according to the scold, is muddled and misbegotten. "Stop reading here" would be the advice of the grumbler.

The scold is spurious. First, Hume made his comment in the context of ridiculing the conviction that reason—a simplistic notion of reason as detached from emotions, passions, and cares—is the watershed for morality. Hume, recognizing that basic values are part of our nature, was unwavering: "reason is and ought only to be the slave of the passions."[3] By *passion*, he meant something more general than *emotion*; he had in mind any practical orientation toward performing an action in the social or physical world.[4] Hume believed that moral behavior, though informed by understanding and reflection, is rooted in a deep, widespread, and enduring social motivation, which he referred to as "the moral sentiment." This is part of our biological nature. Hume, like Aristotle before him and Darwin after him, was every inch a naturalist.

So whence the warning about *ought* and *is*? The answer is that precisely because he was a naturalist, Hume had to make it clear that the sophisticated naturalist has no truck with simple, sloppy inferences going from *what is* to *what ought to be*. He challenged those who took moral understanding to be the preserve of the elite, especially the clergy, who tended to make dimwitted inferences between descriptions and prescriptions.[5] For example, it might be said (my examples, not Hume's), "Husbands are stronger than their wives, so wives ought to obey their husbands," or "We have a tradition that little boys work as chimney sweeps, therefore we ought to have little boys work as chimney sweeps," or "It is natural to hate people who are deformed, therefore it is right to hate people who are deformed." These sorts of inferences are stupid, and precisely because Hume was a naturalist, he wanted to dissociate himself from them and their stupidity.

Hume understood that he needed to have a subtle and sensible account of the complex relationship between moral decisions on the one hand, and the dynamic interaction of mental processes—motivations, thoughts, emotions, memories, and plans—on the other. And to a first approximation, he did. He outlined the importance of pain and pleasure in learning social practices and shaping our passions, of

institutions and customs in providing a framework for stability and prosperity, of reflection and intelligence in revising existing institutions and customs.[6] He understood that passions and motivations, as well as moral principles, can, and often do, conflict with one another, and that there is individual variability in social temperament.

Thus, to continue in the contemporary idiom, the relation between social urges and the social practices that serve well-being is not simple and certainly not syllogistic; finding good solutions to social problems often requires much wisdom, goodwill, negotiation, historical knowledge, and intelligence. Just as Hume said. Naturalism, while shunning stupid inferences, does nevertheless find the roots of morality in how we are, what we care about, and what matters to us—in our nature. Neither supernaturalism (the otherworldly gods), nor some rarefied, unrealistic concept of reason, explains the moral motherboard.[7]

So how did the idea "you cannot derive an *ought* from an *is*" acquire philosophical standing as the "old reliable" smackdown of a naturalistic approach to morality? First, a semantic clarification helps explain the history. *Deriving* a proposition in deductive logic strictly speaking requires a formally valid argument; that is, the conclusion must deductively follow from the premises, with no leeway, no mere high probability (e.g., "All men are mortal, Socrates is a man, so Socrates is mortal"). Assuming the premises are true, the conclusion must be true. Strictly speaking, therefore, one cannot derive (in the sense of *construct a formally valid argument for*) a statement about what ought to be done from a set of facts about what is the case. The other part of the story is that many moral philosophers, especially those following Kant, thought Hume was just plain wrong in his naturalism, and that biology in general has nothing to teach us about morality per se. So they hung naturalism by the heels on Hume's is/ought observation.

But Hume was right to be a naturalist. In a much broader sense of "infer" than *derive* you can infer (*figure out*) what you ought to do, drawing on knowledge, perception, emotions, and understanding, and balancing considerations against each other. We do it constantly, in

both the physical and social worlds. In matters of health, animal hus-
bandry, horticulture, carpentry, education of the young, and a host of
other practical domains, we regularly figure out what we ought to do
based on the facts of the case, and our background understanding. I
have a horrendous toothache? I ought to see a dentist. There is a fire on
the stove? I ought to throw baking soda on it. The bear is on my path? I
ought to walk quietly, humming to myself, in the orthogonal direction.
What gets us around the world is mainly *not* logical deduction (deriva-
tion). By and large, our problem-solving operations—the figuring out
and the reasoning—look like a *constraint satisfaction* process, not like
deduction or the execution of an algorithm. For example, a wolf pack
watches the caribou herd, and needs to select a likely victim—an ani-
mal that is weak, isolated, or young. The pack is very hungry and needs
to be successful, so a lame older animal may be a better choice than a
tiny newborn, but it is more risky; the hunters want to conserve energy,
but acquire a rich energy source; they need to take into account the
location of the river, how they can drive the victim to a waiting pair of
wolves, and so forth. Humans encounter similar problems on a regular
basis—in buying a car, designing a dwelling, moving to a new job,
selecting whether to opt for an aggressive treatment for metastasized
cancer, or hospice care. In any case, that most problem-solving is not
deduction is clear. Most practical and social problems are constraint
satisfaction problems, and our brains often make good decisions in
figuring out some solution.[8] What exactly *constraint satisfaction* is in
neurobiological terms we do not yet understand, but roughly speak-
ing it involves various factors with various weights and probabilities
interacting so as to produce a suitable solution to a question. Not nec-
essarily the best solution, but a suitable solution. The important point
for my project, therefore, is straightforward: that you cannot *derive* an
ought from an *is* has very little bearing so far as in-the-world problem-
solving is concerned.

Brains navigate the *causal* world by recognizing and categoriz-
ing events they need to care about, given how the animal makes a

living—what berries taste good, where juicy termites can be found, how fish can be caught.[9] The hypothesis on offer is that navigation of the *social* world mostly depends on the same neural mechanisms—motivation and drive, reward and prediction, perception and memory, impulse control and decision-making. These same mechanisms can be used to make physical or social decisions; to build world knowledge or social knowledge, such as who is irascible, or when am I expected to share food or defend the group against intruders or back down in a fight.[10]

Social navigation is an instance of causal navigation generally, and shapes itself to the existing ecological conditions. In the social domain, the ecological conditions will include the social behavior of individual group members as well as their cultural practices, some of which get called "moral" or "legal." By and large, humans, like some other highly social mammals, are strongly motivated to be with group members and to share in their practices. Our moral behavior, while more complex than the social behavior of other animals, is similar in that it represents our attempt to manage well in the existing social ecology.

In sum, from the perspective of neuroscience and brain evolution, the routine rejection of scientific approaches to moral behavior based on Hume's warning against deriving *ought* from *is* seems unfortunate, especially as the warning is limited to deductive inferences. The dictum can be set aside for a deeper, albeit programmatic, neurobiological perspective on what reasoning and problem-solving are, how social navigation works, how evaluation is accomplished by nervous systems, and how mammalian brains make decisions.

The truth seems to be that the values rooted in the circuitry for caring—for well-being of self, offspring, mates, kin, and others—shape social reasoning about many issues: conflict resolution, keeping the peace, defense, trade, resource distribution, and many other aspects of social life in all its vast richness. Not only do these values and their material basis constrain social problem-solving, they are at the same time facts that give substance to the processes of figuring out what

to do—facts such as that our children matter to us, and that we care about their well-being; that we care about our clan. Relative to these values, some solutions to social problems are better than others, *as a matter of fact*; relative to these values, practical policy decisions can be negotiated.

The hypothesis on offer is that what we humans call *ethics* or *morality* is a four-dimensional scheme for social behavior that is shaped by interlocking brain processes: (1) *caring* (rooted in attachment to kin and kith and care for their well-being),[11] (2) *recognition of others' psychological states* (rooted in the benefits of predicting the behavior others), (3) *problem-solving in a social context* (e.g., how we should distribute scarce goods, settle land disputes; how we should punish the miscreants), and (4) *learning social practices* (by positive and negative reinforcement, by imitation, by trial and error, by various kinds of conditioning, and by analogy). The simplicity of this framework does not mean its forms, variations, and neural mechanisms are simple. On the contrary, social life is stunningly complex, as is the brain that supports our social lives.

The human capacity for learning and for social problem-solving, constrained by the basic social urges, is the basis for what we commonly think of as social values. To be sure, in different contexts and cultures, particular articulations of those values may have different shapes and shades, even when the underlying social urges are shared. Values are, according to this hypothesis, more fundamental than rules. Various norms governing social life, reinforced by the reward/punishment system, may eventually be articulated and even modified after deliberation, or they may remain as implicit, background knowledge about what "feels right."[12]

Reflecting on the necessities shaping cultures in vastly different conditions and on what social life might have been like for humans living in small groups 250,000 years ago leads us to questions of what distinguishes moral values from other values.[13] I generally shy away from trying to cobble together a precise definition of "moral,"

preferring to acknowledge that there is a spectrum of social behaviors, some of which involve matters of great seriousness, and tend to be called moral, such as enslaving captured prisoners or neglecting children, while others involve matters of more minor moment, such as conventions for behavior at a wedding. The boundaries of the concept "moral," like the boundaries of "house" or "vegetable," are fuzzy even when we can agree on prototypical cases, and this hampers precision in definition.[14] Moral values need not involve rules, though they sometimes do; they need not be explicitly stated, but may be implicitly picked up by children learning to get along in their social world, just as they implicitly pick up how to keep a fire going or how to tend goats.

While acknowledging the central role of cultural beliefs and practices in morality, my aim in this book is to examine the foundations of mammalian sociability in general, and human sociability in particular. I began this project because I wanted to understand what it is about the brains of highly social mammals that enables their sociability and thus to understand what grounds morality. I also wanted to understand variability in social temperament—in the urges to belong, to strongly empathize, and to form strong attachments. Though the approach through the various biological sciences may tell us a lot about the social platform, it is not, by any manner or means, the sum and substance of human morality. Nevertheless, coupled with hypotheses concerning cultural evolution and how culture can change the ecology of a species,[15] the neurobiological perspective may contribute to rounding out the portrait of human moral values that is being pieced together in the behavioral and brain sciences.

My contribution to the science of moral behavior is modest, because many questions in neuroscience and behavioral genetics are still unanswered. It is also very incomplete, because it focuses on the brain, not the recently developed culture in which modern brains live. It is limited because we cannot study the brains or behavior of early humans, nor those of our hominin ancestors.[16] Increasingly, we will learn about the genome of extinct hominins by recovering bits of DNA

from bones, and some information will be garnered thereby. While acknowledging all these limitations, I hope that if my hypothesis is roughly on the right track, it may complement brain and behavioral research.

The core of the biological approach to human morality favored in this book is not new, though my particular way of synthesizing the data and encompassing the relevant philosophical tradition may be. The approach reaches back to Aristotle (384–322 BCE) and the great Chinese philosopher Mencius (fourth century BCE), to those sensible eighteenth-century Scots, David Hume and Adam Smith; it depends enormously on Charles Darwin. Advances in the biological and social sciences have made it possible to explore in earnest the connections between morality and the evolution of the mammalian brain that produced "the family way of life,"[17] and therewith, the wellspring of care and compassion that shapes the moral geography.

Briefly, the strategy for developing the central argument in the book is this: The next chapter will give a bit of background concerning the evolutionary constraints on social and moral behavior. The third chapter goes into detail on the evolution of the mammalian brain and how it supports caring, examining the role of hormones such as oxytocin. The fourth chapter looks more closely at cooperation, especially human cooperation, and data regarding the role of oxytocin in cooperation and trust. The fifth chapter on genes is cautionary, focusing on what is known, and not known, about "genes for" moral modules in the brain. The sixth chapter addresses the social importance of the capacity for attributing mental states, and the possible brain basis for such a capacity. In the seventh chapter, the matter of rules and the role of rules in moral behavior puts the discussion into a more traditional philosophical form. Religion and its relation to morality are the topics of the concluding chapter.