Biodata of **John R. Shook**, author of "***The Design of Morality***."

John R. Shook, **Ph.D**. is a scholar and professor living in Washington, DC. He is Director of Education and Senior Research Fellow of the Center for Inquiry. He also is a visiting assistant professor of science education at the University at Buffalo and associate fellow at the Center for Neurotechnology Studies in the Potomac Institute for Policy Studies in Arlington, Virginia. From 2000 to 2006, he was a professor of philosophy at Oklahoma State University. Among his books are *A Companion to Pragmatism* (2005), *Ectogenesis* (2006), *The Future of Naturalism* (2009), and *The God Debates* (2010).

E-mail: **jshook@pragmatism.org**

L. Swan et al. (eds.), *Origin(s) of Design in Nature*,
Cellular Origin, Life in Extreme Habitats and Astrobiology 23, 161–190
DOI 10.1007/978-94-007-4156-0_9, © Springer Science+Business Media B.V. 2012

THE DESIGN OF MORALITY

JOHN R. SHOOK[1,2]
[1]*University at Buffalo, Buffalo, NY 14260, USA*
[2]*George Mason University, Fairfax, VA, USA*

1. Morality in Nature

Scientific research into our primate cousins and our own human capacities are providing insights into the long history behind the design of morality and our facility with conducting and modifying morality (see, e.g., de Waal, 2009). An account of the origins of human morality must recognize three main phases: a long period of protomorality evolving among the primates; another long period of protomorality growing into morality in the several Homo species as brain size increased; and a relatively brief period down to present times when *Homo sapiens* has been self-consciously modifying morality. Both natural evolution and cultural evolution have been involved in intertwined ways to design morality, and both will continue to shape morality. A naturalistic account of human morality must be both backward-looking and forward-looking. This account looks far back in time when evolution was forming the building blocks of morality and then morality itself without any conscious designing, and then it shows why humans would look forward to take some measure of deliberate control of their inherited morality utilizing only the cognitive resources naturally available to them.

An internally consistent and gradual evolutionary account of these three main phases is needed. Any abrupt break, some strange leap or dramatic shift, presents a severe problem for a naturalistic understanding of the development of morality. For example, if earlier hominid species are thought to have had nothing like morality, only competing in a fierce "selfish" struggle where most kindnesses and cooperations are punished in the long run by natural selection, it becomes hard to explain how *Homo sapiens* would be practicing a far more social morality instead. Some extraneous factor must be abruptly introduced to account for that leap, and speculative theories abound. Perhaps there was a suddenly evolved capacity for universal empathy or for a separate rationality untainted by emotion or ego. Even less naturalistic are proposals about how humans must utilize a spiritual free will uncaused by anything physical or must consult divine revelations descended from above. No great leap is needed, however, if primate behavior displays the building blocks of morality and if the differences between hominid protomorality and human morality are mediated by gradual stages. Nor would there be a forced choice between crediting nature or nurture for morality. The long-standing debate over whether human morality is mostly the product of

natural evolution or cultural evolution presumes a dividing line during our hominid ancestry where morality could be placed primarily on one side or the other. Morality's design is likely more complex than that nature-cultural debate admits.

An objective investigation of morality is needed in the course of pursuing an evolutionary account of human morality. For the sciences to be able to study human morality, it must have some place in the natural world, available for observation and experimentation. Not surprisingly, the most obvious place to investigate human morality is to examine the conduct of natural humans. By regarding humans as entirely natural entities with biological properties and cultural capacities, the sciences have plenty of morality within view for study. Let "moral naturalism" stand for the view that there is a natural phenomenon properly called morality that exists in human societies and that everything about morality's operations or evaluations is open to scientific scrutiny. Moral naturalism holds that human morality can be objectively studied by the several social and natural sciences and that nothing about human morality must elude scientific understanding. This is not the same enterprise as using science to answer our questions about what the morally right thing is to do. As a descriptive effort only, moral naturalism is actually a quite modest and limited enterprise. "What is morality?" is not the same question as "What is morally right?" There is a distinction between studying a practice and having expertise in that practice. For example, we can observe how agriculture works without acquiring farming expertise, and we can observe how people build all kinds of bridges without ourselves having to answer questions about the best way to build a bridge.

Although often taken synonymously, it is useful to distinguish moral naturalism from ethical naturalism. Morality refers to related kinds of actual human conduct, while ethics refers to philosophical questions that arise about morality. Ethics is the philosophical domain that deals with metalevel issues about ways for determining the applicability of moral norms and terms, questions about the appropriate methods for judging and answering moral problems, and concerns over whether one correct morality can be determined. Adding the naturalistic outlook to morality and then to ethics yields two quite distinct fields of study. Moral naturalism takes humans to be doing morality naturally, without any nonnatural features (such as a spiritual soul, free will, or pure reason) involved in the process. Ethical naturalism goes much further than moral naturalism, by regarding all metalevel issues about morality as issues to be resolved naturalistically, and typically includes views that moral rightness and personal goodness are reducible to nonmoral features of nature, that things like moral values and moral facts refer only to natural entities known by science, and that there are true propositions of morality that are made objectively true by nature alone regardless of any human cognition or consensus. Moral naturalism, as the term is used here, is entirely unaffected by the validity or invalidity of ethical naturalism or moral realism. Moral naturalism and ethical naturalism share a respect for science and agree that science should take the lead when investigating morality. Indeed, determining the degree of ethical naturalism's validity largely rests on first carrying out

the program of moral naturalism as far as it can go. No moral naturalist needs to first decide whether any part of ethical naturalism is valid. Indeed, moral naturalism remains useful even if ethical naturalism is invalid. There are good prospects for ethical naturalism all the same. Moral naturalism has made great strides, and there are sound varieties of naturalism capable of grappling with philosophical ethics (see, e.g., Darwall et al., 1997; Foot, 2001; Wong, 2006; Flanagan et al., 2008).

Regardless of ethical naturalism's prospects, we are here only concerned with moral naturalism. Ethical naturalism's interest in discerning moral goodness, value, and truth naturalistically is not shared by moral naturalism. For example, only confusion arises if it is first supposed that there are moral truths, or something like one "true" morality, and that moral naturalism should determine or ground true morality. The project of moral naturalism may not lead to a victory for ethical naturalism. Moral naturalism is compatible in theory with moral relativism, for example – the many varieties of moral societies and their deep disagreements indicate that there may be no way to determine a uniquely correct morality. Still, moral naturalism need not imply any simplistic relativism, since there may yet be better and worse moralities even if there is no perfect morality (see Harris, 2010). Moral naturalism may even reveal that even our "best" human moralities are hardly as good as we think they are because much seemingly moral conduct is actually motivated by selfish concerns, either consciously or unconsciously. All the same, the very fact that humans spend so much time deliberating about good and better moralities, discerning whether each other's conduct is genuinely moral or not, and encouraging each other to be more moral proves that morality is no illusion even if it cannot be so prevalent. Humans put a great deal of effort into morality and, by extension, into ethical inquiry. Moral naturalism's ability to explain how humans now regard morality as variable and modifiable does show why ethical theorizing bothers to seek moral reform and agreement about moral correctness. We envision possible better moralities because we already understand what it is like to redesign inherited moralities.

Because societies do think about what the best morality may be, and whether there ought to be one morality for all, and because societies frequently try to impose one morality on everyone, moral naturalism also gets confused with moral universalism. Despite widely shared hopes, moral naturalism is not a method for determining one uniquely correct morality. Conversely, moral universalism is a bad starting point for undertaking moral naturalism. Those in thrall of one universal morality tend to denigrate as either immoral or nonmoral any mode of human conduct that strays too far from the one correct morality, whatever that may be. Universal moralizers are not alone in this inability to recognize different moral systems. Since we all have been encultured into at least one moral system, any of us can feel susceptible to this inability when we read accounts of the moralities of societies which endorse such things as paternalism or slavery. "That's not any kind of morality," we feel compelled to say, and we may not be assuaged by reminders that those practices are considered moral by

those other societies. Relativism is not intuitive for morally trained people; those other "immoral" societies equally judge our morals be immoral. But scientific objectivity is needed here. Moral naturalism takes its subject matter to be those modes of conduct which peoples themselves have regarded as fulfilling morality, regardless of whether those morals are presently regarded as moral conduct by us. Morals have and will forever vary widely across humanity, and moral naturalism attempts to naturalistically account for all of them.

Moral naturalism limits its interest to the actual moral conduct of humans, but this is a wide field, since all human societies use morality. Unlike many ethical naturalisms and some other ethical theories, moral naturalism does not limit the meaning of "morality" to only some allegedly correct propositions of morality, as if societies unaware of "the" correct morality therefore fail to have any morality at all. Moral naturalism treats morality rather like food production or child rearing: these are things which all human societies do, even if some societies may do them somewhat differently or even better than others. While setting aside the question of one "correct" morality, moral naturalism is primarily concerned with morality as practiced by humans and hominid ancestors, and it uses a definition of morality as humanity now practices it (this definition is coming in the next section). While moral naturalism cannot begin from any single moral system, it must start from a unitary notion of what "morality" is, in order to specify the subject matter for inquiry. Similarly, any study of agriculture starts from a preliminary definition of what "agriculture" consists of, without also premising criteria for some "best" kind of agriculture. It will not serve to leave open what "morality" might be, supposing that empirical inquiry will settle the matter down the road. After all, moral naturalism primarily aims at accounting for what humans currently are capable of doing and not accounting for something we do not do. Although there are researchers inquiring into other species from dolphins to wolves who claim that those species have "moralities" of their own (de Waal, 1996, 2006; Bekoff and Pierce, 2009), their conclusions about nonhuman moralities are just that conclusions about nonhuman moralities. Moral naturalism inquires into the homologous features among human and hominid moralities, regarding them as sharing a common ancestry and remaining grounded in anthropology. It can only recognize near morality or components of human morality in other species. While it may ultimately be a good idea to extend moral relativism to other social mammals and cetaceans, moral naturalism aims at explaining morality as humans have come to practice it.

2. Human Morality

A rough definition of human morality would start from the way that morality is practiced. Morality is naturally embodied in the ways that human individuals voluntarily and habitually conduct themselves in accord with understood norms regulating social interactions and related deeds of social concern, by not only

regulating their own behavior but also by participating in the needed enforcements of moral norms and by teaching these norms and the modes of enforcement to those who need moral education. Morality is primarily designed to regulate social relationships. Moral rules about what a person does in private have their ultimate basis in what society deems as unfit for relationships: disgusting, vulgar, disabling, sacrilegious, or dishonorable deeds that offend society and render a person unfit for some social relationship. Enforcement and education are essential to morality because morality is embodied in the voluntary habits of understood normative conduct. The noticeable way that people frequently avoid or violate norms of conduct reinforces this crucial point about morality: while morality must be to some degree habitual, it must also be voluntary, and hence violable and irregular. By "habitual" we do not mean instinctive, reflexive, or robotic – there is nothing necessary or fixed about learned human habits. Indeed, precisely because acquired habits inculcated by cultural training have only a modest way of guiding our conduct, so much enforcement and education is needed throughout life. On the other side, "habitual" should not be taken to always mean rule following – people can acquire habits by guided imitation and not memorization of express rules, and people usually practice habits without reflecting on any rules governing their habits. Some cultures may get around to expressing and teaching expected moral habits and social roles with explicit rules, and many need not. Self-conscious rule following is not essential to morality but only comprehension of what normally ought to be done in a particular situation.

A morality, like any normative practice, is largely internalized. When people are conducting themselves morally, they are following moral norms not because they feel a sudden urge to be kind, or they are forced or coerced to do so, or because they strategically regard obedience as simply a means to obtain what they really want, but rather because habitual respect for moral norms and other people are among their own important motivating values. Neither sympathy to needs, obedience to commands, nor compliance with expectations is sufficient to constitute morality (although those three factors can enhance moral conduct). A person is not being moral by feeling motivated to help because a sudden discomfort of sympathy or pity has been aroused. A person is not being moral simply by obeying a command because it is backed by threats of punishment that she wants to avoid. A person is not being moral by complying with a rule of conduct because compliance is the best way to get what he wants anyways. A person is behaving morally when they willingly conform to a moral norm because this person's respect for that norm is a sufficient and effective motivation to habitually want to conform.

Moralists frequently demand that genuine morality must spring from a purely altruistic desire to help another regardless of any estimation of duty. Perhaps the idea that morality should be like the characteristic altruism of close kinship has perpetuated this intense moralism. Finding so little angelic purity across humanity, disappointed moralists are heard to complain that people have little or no morality, but such cynicism arises from looking in the wrong place for

morality. Philosophical ethicists frequently demand that genuine morality occurs when that person's respect for moral obligations provides an overriding and compelling reason for complying regardless of that person's emotions, desires, or values. But we are not undertaking philosophical ethics here, either. Moral naturalism need not postulate anything like a detached rationality capable of dictating conduct quite apart from emotions, desires, or values. This is fortunate, since there may be no such detached thing in human psychology (see, e.g., Gazzaniga, 2005; Greene, 2008). Nor is moral naturalism premised on moral realism or on any claims that truth or rightness attaches to the moral judgments people make. Moral naturalism is unaffected by the alleged prevalence of vast moral error (as claimed by many such as Joyce, 2001; Lillehammer, 2003), and morality has not been "debunked," so moral naturalism still has a subject matter. Whether morality is actually what people suppose it to be is a concern for philosophical ethics, not moral naturalism. Despite the fond dreams of some moralists or ethicists, neither pious altruism, pure reason, nor perfect judgment has been powering human morality, even if we may presently want to modify morality to become more altruistic or rational.

Moral conduct, when it occurs, is primarily motivated by a perceived duty. There may be other motivations to conform as well: nonconformity may bring unwanted punishment; conformity may be a means to get what one really wants; or conformity may assuage one's uncomfortable emotions. However, when a person sufficiently respects a norm, that person conforms even where nonconformity may bring no punishment or personal benefits, and even if no compassionate emotions are dominant. There are degrees of respect for moral norms, and enforcement and education are common means to increase respect. Enforcement and education rely on the deeper morality-building motivations of caring for others, avoiding harms from others, or getting benefits from others, but encultured morality results in motivating habits not reducible to any combination of these more basic and evolutionarily older motivators. Because human societies can promulgate a wide variety of moral habits, we may say that in a sense, morality is socially conventional, but only in one sense. While morality persists in human culture largely because humans do the training, that does not mean that humans must regard their moralities as conventional. Indeed, many societies teach that their own morality is the only morality, and some additionally teach that their morality is grounded on nonhuman matters like nature's ways or a god's wishes. A well-trained moral individual is not likely to regard moral duty as based merely on what society wants – the moral individual is more likely to regard what is morally right as enjoying a foundation independent from humanity. Similarly, although the practices of mathematics persist because human societies promulgate them, mathematics practitioners are unlikely to regard the truths of mathematics as dependent on what society happens to promulgate. Socially designed morality is functioning well for people when they do not regard it merely as locally and conventionally designed. We must not make too much of this looming paradox when we are pursuing moral naturalism. It is a peculiar problem for ethics, and

not moral naturalism, to help reconcile the seeming independence of general moral duty with its actual dependence on local moral education.

Morality is designed to function somewhat differently from other socially normative practices. In human societies, morality can be distinguished from two similar social practices concerning relationships that extend beyond the range of the family: the norms of political laws on the one side and the norms of etiquette on the other. Political laws backed by government force can secure widespread and uniform obedience from the people, but entirely voluntary respect for those laws may be weak or far from universal. Laws backed by effective threats do their proper work of preventing harms and violence by appealing to an individual's basic desires to avoid public shame and harsh punishment. Civil etiquette is commonly quite voluntary, but it can vary so widely among individuals in the same society, and so irregularly enforced by sanctions, that specific norms accepted by all cannot be clearly identified. Nevertheless, norms of etiquette do their proper work of promoting aid and altruism by appealing to an individual's basic capacities for feeling sympathy and compassion toward others. Research on primates indicates that a protomoral sense of compassionate altruism and respectful fairness can occasionally be observed in their social behavior, approximating what is labeled here as "etiquette" (not in the sense of refined manners, of course, but just the simple gestures of nice and fair treatment).

Morality is in the middle ground between law and etiquette and overlaps them on each side – it concerns specifiable norms about social relationships and interactions for the entire society to willingly conform and mutually enforce. In every human society, such norms are evident and powerful, although their specific scope and content varies widely across cultures. A type of social interaction dictated by morality in one culture may be left up to etiquette in another; what is assigned to law to enforce in one culture may be left up to morality in another; and some societies may not regulate some kinds of conduct which other societies heavily regulate using morality and law. However, morality has a distinctive role in every society: it concerns those norms in a society where that society regards them as both universally applicable and universally worthy of sufficient respect. Society demonstrates its regard for morality by expecting voluntary conformity to its norms, expecting people to help enforce conformity where needed, and expecting people to help instruct the young to acquire moral norms. Where society encounters individuals unable or unwilling to conform morally, even if they conform obediently or self-servingly, that society expends efforts to reform that person's attitudes and habits, or failing in that effort, reclassifies them into some subnormal status (e.g., with unreformed criminals, the mentally ill, or the cognitively disabled).

Just as morality, while having universal applicability within a society, can vary in scope and content from society to society, morality can vary in effectiveness. If it were somehow impossible for morality to vary in its effective impact on the lives of society's members, they would not even think to try to modify it. Since morality obviously can and does have varying noticeable impacts on different

individuals, so where there is sufficient intelligence, morality can therefore become an object of interest and an object of manipulation.

The opportunities for deliberate modification of morality are everywhere, since individuals can acquire some intelligent control over their conduct, even much of their habitual conduct, if they can consider their conduct as something controllable and modifiable over time. The story of the evolution of culture is essentially the story of the increased capacity for humans to regard some of their habitual conduct as modifiable with attention and practice and as teachable through instruction. The story of the evolution of morality, as a mode of enculturation, is essentially the story of the increased capacity for hominids to regard and enforce some social norms as worthy of everyone's willing conformity. As objects of intelligent attention in their own right, humans then gradually came to regard such moral norms as deliberately modifiable and proceeded to experimentally redesign the many moralities now embodied in diverse human cultures. By analogy, early hominids developed habitual tool use over two million years ago, but few modifications to choppers, flakes, and blades occurred until brain size had dramatically increased; the immense proliferation and complexity of tools associated with *Homo sapiens* indicate how hominids and humans gradually took deliberate control over experimental tool construction.

3. The Function of Morality

Humans would not experimentally redesign morality unless it came to be viewed as something not just modifiable but as modifiable for serving some end. An experimental modification to something, as opposed to an accidental, sporting, or aesthetic modification, treats it as a means for achieving envisioned consequences. What is it like to regard morality as a means functioning to serve ends? What would be the function of morality? We have located morality among the modes of human conduct, where individuals willingly regulate their social interactions out of respect for its norms and they expect everyone to do likewise. This descriptive view of morality omits its functionality: why would humans have morality? Couldn't human societies do well with just familial altruism and social etiquette, or some combination of familial love, etiquette, and law?

Morality supplies something that neither familial love, etiquette, nor law can provide. Morality permits standardized modes of interactions which each individual can reliably expect from all others under conditions when kinship is absent, etiquette is doubtful, and punishment is uncertain, unwise, or too costly. Etiquette is not standardized and not evenly enforced; indeed, norms of etiquette are precisely those optional norms which lack overriding respect and little punishment if any is attached to their enforcement. (That is why reciprocal altruism cannot be the essence of morality but only a display of optional etiquette – see Tullberg 2004 on differences between altruism and reciprocity.) The most important norms of etiquette overlap with the minor norms of morality, and in modern

human societies that range of normative conduct makes up what we call civility. On the other side, where a society uses law, the most important codes of morality overlap with much of law, and that range of normative conduct presently makes up what can be called good citizenship. Law promises standardized modes of interactions and high probabilities of punishment, but enforcing law has many costs because it is very intrusive on individuals and requires large resources for adjudication and punishment; that is why societies which use laws are regulating only the most important kinds of social interactions. Morality serves to regulate conduct across a broad array of human interactions where norms must be both universally respected and efficiently enforced.

What sorts of social interactions would benefit from something like morality? The obvious kinds of interactions are cooperations. Moral norms, because they regulate everyone's conduct in an efficiently uniform manner, are highly useful for promoting cooperation among all members of society. Where the more basic components of morality are already in place, especially the norms of etiquette encouraging mutual assistance and fair treatment, stable patterns of mutually beneficial cooperation can emerge and grow. If individuals can be confident that mutual assistance will reliably yield sufficient benefits, without worry that unfair treatment might occur, they will naturally undertake cooperative projects with some frequency. In short, the original effects of occasional friendly etiquette can magnify into the repetitive trusting cooperation that can deliver even higher benefits for individuals in intensely social societies. Indeed, a highly social society is precisely that society in which both simple etiquette and complex cooperation are regularly occurring to the high benefit of all members across lifetimes and over generations.

The mutual helpfulness and fair cooperation permitted by basic etiquette can yet remain unstable and less frequent compared to other strategic modes of social interaction going on within a society. To become the dominant mode of social interaction, robust cooperation across societies requires more than just etiquette. The additional assistance comes from morality's universal and stricter obligations. The deliberate invention and design of law was a further extension of this same process in highly complex human societies, increasing the benefits of cooperative social environments by applying more costly regulation to the most important behaviors threatening the proper functioning of a civil society. The continuities between stages of moral development, proposed at the outset of our expectations for moral naturalism, can be observed in theory and fact. If etiquette had utility for small groups, morality would evolve that utility to larger groups, and law extends that normative development to fully complex societies, making higher civilization possible. A society concurrently utilizing all three modes of etiquette, morality, and law would then maximize efficient continuity: making everything a matter of law would be too costly, as would folding all of etiquette into morality. It may be theorized that an optimal society, therefore, would attempt to match its expenditures for normative implementation to the social significance of the behavior regulated. We do observe many modern human

societies displaying such effort to efficiently utilize etiquette, morality, and law concurrently, although societies design their own distinctive modes of assigning expected conduct to each category.

4. Morality and Cooperation

Intuitively, cooperation and morality are a good functional match. However, we must not hastily assume that morality only exists where intense cooperation is ongoing, and we must not assume that cooperation requires morality. Species can evolve intense forms of social cooperation without any morality or even social intelligence (take ants for example) because such cooperation can be sustained by close kinship alone. For its part, morality could theoretically serve other functions besides cooperation. There are competitive "zero-sum" games of winners and losers, what can be called "win-lose" games, that can be better sustained if some moral norms are added to the rules of the game. However, the impressive value of morality is revealed when it is added to cooperative "win-win" or "non-zero-sum" games of mutual benefit. Indeed, it is hard to imagine how any sustained forms of intense cooperation across large societies would last without support from moral norms. All the same, the question stands: could non-zero-sum cooperation be worth it, relatively speaking? Numerous studies of forms of reciprocity generally suggest that this is the case (see, e.g., Trivers, 1971, 1985; Hirshleifer and Martinez Coll, 1988; Sober and Wilson, 1998; Henrich et al., 2001; Sachs et al., 2004; Taylor and Nowak, 2007).

It is widely agreed that, in theory, non-zero-sum cooperation tends to delivers greater overall benefits to most individuals over time in a complex society than any combination of competitive zero-sum games, provided that individuals have cognitive skills sufficient for conducting such social cooperation. A resulting overall advantage of cooperation over competition would practically explain why larger groups of increasingly intelligent hominids and humans evolved to have greater reliance on non-zero-sum cooperation than all other zero-sum interactions combined. Cooperation for common and mutually beneficial ends is a pervasive and important feature of any human society; indeed, across several hominid species, social cooperation largely determines the welfare of any of its members. For modern humans especially, avoiding social cooperation is the path to death; even a "self-sufficient" hermit or a combative aggressor was taught survival and fighting skills by others.

Where does the role of morality enter? The modest amount of social cooperation that gradually arose in early hominid species would have benefitted greatly from simple moral norms, over and above the contributions to cooperation made by the sentiments of sympathy, compassion, and altruism. Those crucial sentiments secure the bonds of commonality among closely related kin and a few caring friends, but they do not function well for impartial extension to everyone. The value of morality only increased as *Homo sapiens* gradually lived in larger

tribal units and invented more complex forms of social cooperation occurring across the boundaries of close family and friends. Where the emotions guiding family commonality are absent, a different kind of cooperative relationship with acquaintances can develop instead, one based on reciprocity. Given the above sketch of what morality is and how it basically functions, we can understand why morality is well-suited to facilitating widespread cooperation for high mutual benefit. The universality and overriding respect for moral norms is a good functional match with the widespread and repetitive modes of cooperation going on within social groups. Reliance on etiquette does not vanish as reciprocal cooperation increases, but its unsteady and varying force would enjoy a dramatic supplementation with morality. In fact, etiquette and morality would reenforce each other among the most cooperative members of a society; that is why distinguishing merely compassionate acts from genuinely moral conduct is no simple matter. (Similarly, when modern humans invented political law, it hardly replaced etiquette and morality, but supplemented them in an integrated fashion, so that refraining from murder is simultaneously legal, moral, and nice.)

Where any two members of a group can both have some assurance that significant mutual benefit is possible and no unfair harm is forthcoming, they have a greater rational incentive to engage in reciprocity cooperation. Core moral norms helpful for such cooperation would naturally include prescriptions against coercion or harms by physical domination, deception, cheating, stealing, and unfair treatment. Respect for morality provides needed assurances for all parties, and the value of that assurance increases with the size of the social group. In very small groups where everyone is closely related and quite familiar to each other, instinctive emotions of sympathy and compassion could be strong enough to ensure care and cooperation, so familial love and kind etiquette is usually sufficiently and no morality is needed. But in larger groups, where individuals will encounter distant relations (or nonfamily as well), morality adds its distinctive service to increasing cooperation over and above the bonds of family and vagaries of etiquette. In such societies, where individuals could benefit from cooperation with others known primarily by reputation, moral norms increase the chances of successful cooperation. In effect, morality would be displayed in situations where two or more individuals know each other primarily by reputation, they have repeated opportunities to mutually benefit from cooperation, and they conform to moral norms about cooperating fairly and not harming each other in the process.

This view of morality's functioning presupposes that individuals have enough cognitive resources to absorb and understand prevailing moral norms in their society, to know how to strategically comply or ignore those moral norms, and to judge the other members' reputations for complying with those norms as well. There is no well-established theory about when early hominids developed all of these cognitive resources and began doing what we know as morality. Perhaps our primate cousins are capable of occasionally performing very simple versions of morality. But it may prove difficult to determine whether their conformity to norms is due to sympathetic feelings, strategic aims, or actual respect for the

norms themselves (just as judging motives of fellow humans is not easy). In any case, as hominids gradually came to rely on social strategies demanding intense daily cooperation, it is reasonable to suppose that the basic primate toolkit for protomorality underwent development too. That development led to us: we can observe how the intelligence of *Homo sapiens* is highly developed for tracking the conduct and reputations of many individuals and for enforcing and teaching morality right along with the other aspects of complex culture.

When morality is supplemented into the manner in which individuals cooperate, this moral cooperation is dramatically enhanced in two ways: they will cooperate more frequently, and their cooperation will be much more efficient. Where cooperation is conducted morally, the participants will generally display a high level of cooperativeness with most if not all other members of the society. Specifically, each member will tend to be willing to cooperate when there is an opportunity, and each member will take an opportunity to cooperate with most or all other members of the society as situations warrant. In simplistic terms, where morality is robust, we would expect to observe lots of niceness and very little choosiness. On an emotional level, robust morality is experienced as trust: a felt confidence that reciprocity interactions with another will be reliably safe, beneficial, and fair.

We should not overrationalize the gradual emergence of a rational morality within social groups. It is unnecessary to depict hominids as cold calculators thoughtfully rationalizing cooperation. The cognitive capacities needed for morality's emergence would not need to be more sophisticated than the "fast and frugal heuristics" of severely bounded rationality (see Gigerenzer and Selten, 2001). Like so many other evolved features of hominid brains, the evolving emotions do most of the work of social intelligence anyways, and humans have inherited and use them (see Haidt, 2003; Nichols, 2004). Individuals growing up in a relatively cooperative society will be guided by the feelings of enjoying mutual trust. Nor is necessary to depict early hominids or *Homo sapiens* as utilitarians who regard morality as a means to enhancing social welfare. Actually, individuals committed to morality will instead generally regard morality as binding regardless of calculations of social welfare. On a personal level, a moral person will try to be virtuous on a daily basis, and a group of moral people will regard each other as dutiful agents. Perhaps only when a society's morality is under widespread dispute or under discussion as a problem in itself would people think to inspect morality in regard to its service to society. Whether virtue theory, deontology, or utilitarianism has greater merits is an issue for philosophical ethics, not moral naturalism. For its part, moral naturalism can at least account for the distinctive ways that morally trained people would regard how morality works. Moral psychology has the unenviable task of sorting out the actual motivators for moral judgment from the stated justifications that people will supply when asked (see, e.g., Doris et al., 2010; Brinkmann, 2011).

So far, a naturalistic explanation involving both natural and cultural evolution for morality's enhancement to routine and widespread cooperation seems possible.

However, as already noted, there are many other strategies for conducting social interactions besides morality. Even if we might see how morality can enhance cooperation, individuals can still benefit from competitive, neglectful, unfair, or harmful conduct toward others too. Why would morality come to have any large role to play in hominid societies? Perhaps morality has long been just one optional mode of social conduct among many others, and individuals could vary widely in the extent to which they rely on morality over other strategies. Perhaps a society that does use much morality could be successfully "infiltrated" by individuals using more selfish strategies. Is there good reason to conclude that individuals in societies using lots of moral cooperation enjoy enough benefits to survive or to conclude that societies using moral cooperation might just as well drift away from morality as time goes by?

There is no point to accounting for the widespread reliance on morality among contemporary humans with an evolutionary story of hominid development unless moral cooperation can work well, at least better than zero-sum interactions. It is unnecessary to think that the proliferation and supremacy of nonzero cooperation was in some sense evolutionarily inevitable (as suggested in Wright, 2000), but there are good reasons to think that plenty of non-zero-sum cooperations would be stumbled upon and repeated by sufficiently intelligent species, such as our primate and hominid ancestors, giving rise to culture (a survey of reasons is given by Boyd and Richerson, 2005; Nowak, 2011). Are there good reasons to think that moral cooperation delivers generally high benefits to the members of a society and delivers results more beneficial to all individuals within that society than any other strategy? Further study is needed to analyze the benefits of moral cooperation in a society.

A commonly used tool for analyzing the supposed benefits of reciprocity cooperation is the prisoner's dilemma game. Computer modeling of groups utilizing various strategies shows that a small number of simple strategies, such as "Tit For Tat," tend to garner the most benefits (though there is no provably optimal strategy for indefinitely extended iterations). However, the prisoner's dilemma game models a very specific kind of basic social interaction, which is too simplistic to bring morality into full view. What is needed is a revised type of prisoner's dilemma game that models the more complex kind of social interaction where emerging morality would be occurring.

5. The "Indirect Reciprocity" Prisoner's Dilemma Game

The prisoner's dilemma is a way to model simple interactions between individual agents where there is the possibility of reciprocity, cooperation, and betrayal. This section describes a computer program which models a tournament of players taking part in what I have termed the "indirect reciprocity" version of the prisoner's dilemma game, or the IRPD game. I first presented the IRPD game at the Seventh Annual Conference on Computers and Philosophy at the University of Central

Florida in August 1992, and a more sophisticated version was presented at Oklahoma State University in 2003. The basic inspiration for this project comes from Robert Axelrod's (1984) ground-breaking exploration of the prisoner's dilemma. The study of the prisoner's dilemma has been widely perceived as an opportunity to gain insights into the nature and origins of morality. A full explanation of the prisoner's dilemma and its importance is found in Axelrod's book and subsequent books on its analysis (see, e.g., Colman, 1995; Barash, 2003).

The essential point is that while the prisoner's dilemma (PD) is a single interaction, the prisoner's dilemma game is a long series of such dilemma interactions, where neither player knows when the series will end. A strategy is a method of playing this PD game. The PD by itself has one optimal strategy regardless of the opponent's strategy: defect. The PD game by contrast need not have one optimal strategy, for reasons involving the uncertainty regarding when the last interaction will occur and, if groups of players are interacting, the variability of the strategies involved. The study of the PD game primarily involves the examination of the performance of the possible strategies available to the players, relative to the size and characteristics of the rest of the group.

Axelrod studied how various strategies, embodied in computer programs, performed as they competed in round-robin tournaments. His use of the PD game involves players in a round-robin tournament who use strategies to try to obtain the most points from PD games with the other players. The crucial aspect to Axelrod's methodology is that the strategies permitted by Axelrod can only exhibit direct reciprocity. This direct reciprocity is characterized by the information upon which a player has to base the decision whether to cooperate or defect in a dilemma: the only information available to a player is the history of interactions he has had with the other player. This explains the results of Axelrod's tournament; the winning player's strategy, Tit For Tat, is not nearly as sophisticated as most of the forms which moral behavior can take. Is it possible that Tit For Tat is a morality? As far as familiar codes of morality go, it does resemble the ancient "eye for an eye" rule of retribution: what you have done to me, I will do to you. The important question is whether the methodology of using the PD game can be altered so as to permit more complicated forms of behavior. The answer lies in understanding reciprocity.

There are two kinds of reciprocity: direct and indirect. Direct reciprocity is of the form: A helps B, B helps A. Indirect reciprocity is of the form: A helps B, B helps C, C helps A; or A helps B, C who is observing, later helps A, A helps C. Alexander (1987) and Boyd and Richerson (1989) have emphasized how the more complex behaviors associated with morality are mostly dependent upon the ability to use strategies that exhibit indirect reciprocity instead. Indirect reciprocity requires an additional piece of information: the history of the opponent's interactions with the rest of the players in the tournament. After a discussion of direct and indirect reciprocity, the basic scheme of the "indirect reciprocity" prisoner's dilemma game computer program is laid out, and preliminary results from this program are described. The IRPD game shows how genuine moral cooperation

can be highly beneficial and evolutionarily stable. The following brief account of the IRPD computer program describes its capacity to formulate strategies, group various strategies together, and have them interact in tournaments in a way similar to Axelrod's. The key departure from Axelrod's work lies in the difference between direct and indirect reciprocity.

In the PD game, help comes in the form of cooperation. Direct reciprocity only requires a player to have access to information on an individual level: a record of how the other player has treated one in past dilemmas. The modeling of indirect reciprocity requires additional information: knowledge of the other player's interactions with the rest of the group. Additionally, the modeling must permit a player to make a choice as to whether she will engage another player in a prisoner's dilemma interaction. No such choice is allowed in the direct reciprocity model.

For our purposes designing a basic IRPD game, we shall assume that every player has complete and reliable information; other models of the IRPD game need not do so. We shall also assume that no player will err in making decisions based upon the information; other models may include the possibility of judgment error. On this foundation, we can proceed to construct a prisoner's dilemma model of indirect reciprocity. What follows is but one way to model indirect reciprocity, using discrete, agent-based modeling that tracks the performance of identifiable agents as they interact or avoid each other over time. Aggregative modeling of populations undergoing interactions involving indirect reciprocity has already been attempted (see, e.g., Nowak and Sigmund, 1998; Panchanathan and Boyd, 2004). However, aggregative population dynamics assumes that every individual has an equally likely chance of interacting with every other individual. Because indirect reciprocity permits an individual to abstain from an interaction, as well as to engage in an interaction, full indirect reciprocity is imperfectly modeled by aggregative modeling (see Hauert et al., 2008 for discussion of this point). Indirect reciprocity also works best for finite populations in which members possess some information about everyone's reputation for cooperating or not. Without an estimate of reputation, indirect reciprocity gains little for cooperators, while much knowledge of reputations can produce stable cooperative societies (Milinski et al., 2002; Nowak and Sigmund, 2005).

The IRPD game described here is probably one of the simplest models for agent-based modeling of strategies for undertaking indirect reciprocity with all other social members in a finite population that include the basic features of reputation, avoidance, cooperativeness, niceness, and choosiness. We shall first look at the structure of a strategy which can enable a player to engage in indirect reciprocity. A look at the structure of the round-robin tournament follows.

As we know, indirect reciprocity requires that each player should have the option of whether to have a PD interaction with another player. The judgment that entering into a dilemma with a particular opponent is not in a player's best interests will result in a "shun"; the player will opt out of a PD interaction with the other. (In the basic kind of IRPD game described here, a "shun" is the only mode of "punishment" exacted by a strategy.) In the IRPD game, this judgment

will be made upon the information of how often the other player cooperates in all the games he has played. When a player has decided to enter into a PD interaction, the strategy must direct the move to be made. Accordingly, a strategy will require access to three kinds of information: (a) how often the player should cooperate, (b) how often the opponent has cooperated in all of its interactions with all other players in the past, and (c) what standard the player should use to evaluate other strategies.

The first kind of information is placed in the form of a number ranging from 0 to 1, representing what the frequency of the player's choice to cooperate in an interaction will be. 1 = always cooperates, 0 = always defects, .5 = there is a 50% chance of cooperation, etc. The name for this number is niceness. For the purpose of modeling agents with limited cognitive and deliberative resources, we may regard a player's niceness more as a habit rather than a thoughtful choice upon each interaction.

The second kind of information shall be just the opponent's niceness. This represents each tournament as taking place in the course of the life of a society, after each player has established a reputation among the group. Again, for the purpose of modeling agents with limited cognitive and deliberative resources, we may regard a player's understanding of an opponent's niceness not as the recollection of all the outcomes of the opponent's past interactions, or the recollection of what others have gossiped (no presumption of advanced language use is made here), but rather just as an understanding of the opponent's reputation as far as a player can tell from some observation.

The third kind of information shall again be placed in the form of a number ranging from 0 to 1, representing the standard used to evaluate the other player's niceness in order to make the decision upon whether to interact in a dilemma with the other player (1 = will only choose to play will those having niceness of 1, 0 = will play with anybody, .5 = will play with only those having a niceness of .5 or higher, etc.). The name for this number is choosiness, and again this represents only a habitual inclination of the player and not a series of reflective deliberations.

Since a player's niceness and the choosiness do not change for a strategy during a tournament, we can simply refer to these two qualities as a strategy. In this way, we can say that the entire possible range of strategies can be represented by the points on a one unit by one unit graph.

To get an IRPD tournament started, the decision as to which strategies shall play in a tournament must first be made. The program I wrote permitted the user to form the group either by setting each strategy's niceness and choosiness, or to choose a preset group by selecting a size equal to a perfect square (4, 9, 16, 25, etc.) The preset groups have an even distribution of qualities: with four members, the strategies are 1,1 1,0 0,1 0,0. With nine members, the preset group looks like: 1,1 1,.5 1,0 .5,1 .5,.5 .5,0 0,1 0,.5, 0,0. With groups of 400 and more, the tournament can be very inclusive. Each of the preset groups will have an average niceness and choosiness of .500. Other kinds of preset groups are possible as well, with differing distributions, but await a future version of my program.

A round-robin tournament for the group is scheduled as follows. In one "round," a designated strategy will meet (have the opportunity to interact) each of the other strategies consecutively. In the span of one "season," a number of rounds take place, equal to the number of the strategies in the group, so that each of the strategies will take the role of the designated strategy in one season. Each strategy will thus meet every other strategy twice in one season. As the computer program proceeds through the schedule, for each scheduled meeting allows the two strategies to first decide if the PD interaction should take place. If either player declines because the opponent's niceness is not great enough (does not equal or exceed the player's choosiness), no PD interaction takes place and the programs proceed to the next scheduled meeting. If both do choose to interact, then each strategy decides whether to cooperate or defect in the PD interaction. This decision will be based on the strategy's niceness, by finding whether a cooperation would bring its overall record (ratio of cooperations to total games played) closer to its niceness or not. After each has chosen, the program determines the outcome of the game by assigning points accordingly (e.g., in the standard assignment, 3 points go to each if they both cooperated, 1 point goes to each if they both defected, or 5 points goes to the defector and none to the cooperator; the points awarded can be preset by the user). The computer program saves a database of each player's cumulative points, for reporting at any selected stage of a tournament.

To summarize the program:

1. The number of players and their assigned strategies are chosen by the user.
2. The tournament length (the number of seasons to play) is chosen by the user.
3. The ordering of the group is scrambled.
4. In each season, the necessary number of rounds occurs.
   A. Inw each round, the meetings are assigned.
   B. For each scheduled meeting:
      a. The program determines if the two decide to interact.
      b. If they will interact, it finds each strategy's chosen move.
      c. The outcome of the game is found and points are awarded.
   C. At the end of the season, the ordering of the group is rescrambled.
5. Each strategy's cumulative points are recorded for later reporting.

There are some ways to complicate the IRPD game so that the player's environment and playing conditions more closely approximate the "real world." In the real world, organisms have to "pay" the costs of living from day to day, they will someday die from natural causes, they must reproduce to continue their genetic line, their offspring can have mutated genetic codes, and so forth. In the game, a preset amount of points can be subtracted from the player's total at each scheduled meeting (this is called the "meeting cost"), so that death could occur (total points dropping to zero or below). When a meeting cost is exacted, all strategies begin with an allowance of points equal to 50 times the meeting cost. This amount is not figured into the final results and succeeds in preventing a strategy's

chance bad luck in meeting assignments in the early stages from prematurely killing it. In effect, this tries to compensate for the mere pseudorandomness of the computer and simulates the natural fact that typically organisms are not born (nor enter adulthood) on the very edge of death; they have some (however small) stored resources.

The players who are able to accumulate points at a faster rate will tend to survive, and then the effects on a player of competing in an evolving group can be studied. The players can be given a preset lifespan, in terms of a maximum number of scheduled meetings, after which "death" occurs. In the program, the timing of death is not the same for all: the probability of death for a strategy begins some meetings before the preset lifespan and steadily increases each meeting afterward, so that the actual death can occur in a range around the lifespan.

In more sophisticated versions of the computer program, players can use points to have offspring and use up points just to keep living. To "give birth," a player can be permitted to create a clone of itself after it has accumulated a preset amount of points. The new player has the identical strategy as its "parent" and is entered into the tournament, and some amount of points are subtracted from the parent's total. However, giving birth will not unduly harm a strategy. It is only set back to its original amount of points and keeps its margin of safety from lack of randomness if there is a meeting cost in a tournament. The result is that a more successful player will have a higher birthrate. Also, at the cloning, a mutation can be permitted, to change a new player's strategy so that it differs from the parent's. In addition, the meeting cost can be variabilized according to a change in the size of the group, so that meeting cost rises proportionately to the group size if the group size should rise through births. This models a realistic aspect of a group living in an environment with limited resources, where a rise in population makes it harder for all to survive.

When these complications are put together, natural evolution is modeled. The first generation of strategies can compete, give "birth," and die, giving rise to the second generation, which in turn repeats the cycle. In this way, the evolutionary success of a strategy can be measured. Also, the composition of a group of strategies can be evaluated over many generations for survival (does the group grow, or does it die off) and for total overall success (the cumulative amounts of points of all living strategies).

6. The IRPD Game Results

Some basic results have been confirmed through many hundreds of tournament experiments, and they are consistent and very striking. When there is no meeting cost exacted, the winner always has a very high niceness and a very low choosiness. In fact, the ranking of the strategies by points earned consistently follows the degree to which a strategy's niceness approaches 1 and choosiness approaches 0. The worst finishers are quite nasty and very choosy. For example, the strategy (1, 0)

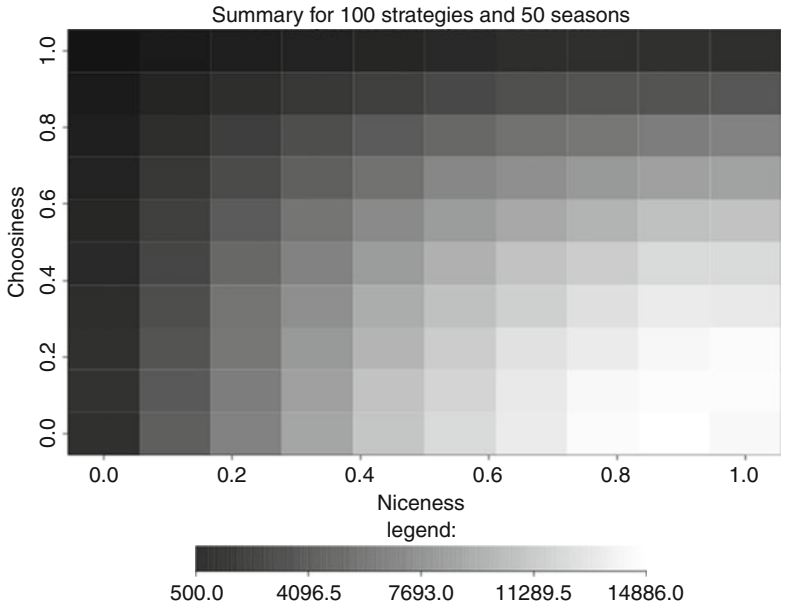

Figure 1. One hundred strategies after 50 seasons. The worst strategies accumulate less than 4,000 points.

nearly always finishes in the top 5%, regardless of the size of the preset group, and the strategy (0, 1) finishes in the bottom 5%. I have been able to confirm this for groups containing up to several thousand strategies. Figures 1 and 2 show arrays of the points accumulated by strategies in groups of 100 and 224 who have played typical tournaments lasting 50 seasons.

If there is a meeting cost exacted from all strategies, four major phenomena occur. First, some strategies will die at some point in the tournament because they fail to accumulate points fast enough to cover the subtracted meeting costs. Without exception, the nastier and choosier strategies die first, so the meeting cost results in a gradual rise in the average Niceness and a drop in the average choosiness until no more strategies die. This should be expected, given knowledge of the results of tournaments lacking a meeting cost in which the nastier and choosier strategies garnered fewer points (gained points at a slower rate) than the rest. In tournaments with a meeting cost, these strategies' points would drop closer to zero, and those whose rate of points per meeting remained less that the meeting cost rate eventually die. The second phenomenon is observed when the meeting cost is varied. There is a direct relationship between the size of the meeting cost and the group's average niceness; an inverse relationship holds between the size of the meeting cost and the group's average choosiness. For example, in a tournament in which the meeting cost is set at .5 point per meeting, the average niceness increases to around .60 and the average choosiness drops to

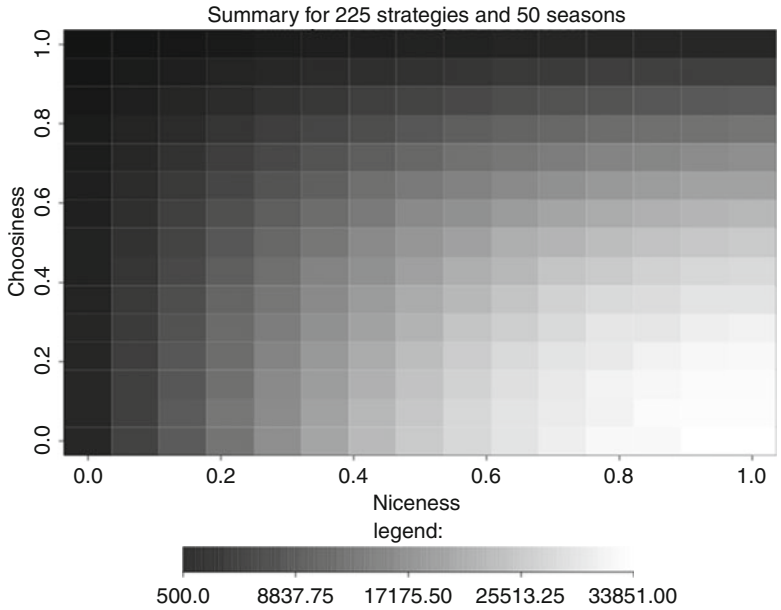

Figure 2. Two hundred and twenty-five players and 50 seasons. No matter the number of strategies, the same strategies perform the worst.

around .40. If the meeting cost is 1.0 point per meeting, the average niceness is around .66 and the average choosiness is around .34. Figures 3 and 4 show an array of 100 strategies playing tournaments of 500 and 1,000 seasons with a meeting cost of .5 point.

If the meeting cost is 1.0 point per meeting, the average niceness rises to around .66 and the average choosiness drops further to around .34. Figures 5 and 6 show an array of 100 strategies playing tournaments of 500 and 1,000 seasons with a meeting cost of 1.0 point. Keep in mind that these are summaries of many independent runs of tournaments so they will display inconsistencies among them. For example, Fig. 5 shows how the strategy (.5, .1) had died in its tournament of 500 seasons, while Fig. 6 shows how the same strategy managed to survive to the 1,000 season mark in another tournament.

This variable meeting cost trend toward higher niceness and lower choosiness does not seem to persist indefinitely, however, as very high meeting costs stabilize the niceness and choosiness of the few surviving strategies in the vicinity of .50 and .40. This experimental model result conveniently corresponds to the familiar resurgence of uncooperative selfishness among primates and humans under very stressful conditions.

The third phenomenon is also seen when the meeting cost is varied: the leaders in these tournaments are not as nice nor as indifferent as those from tournaments lacking a meeting cost. As the meeting cost rises and the group becomes

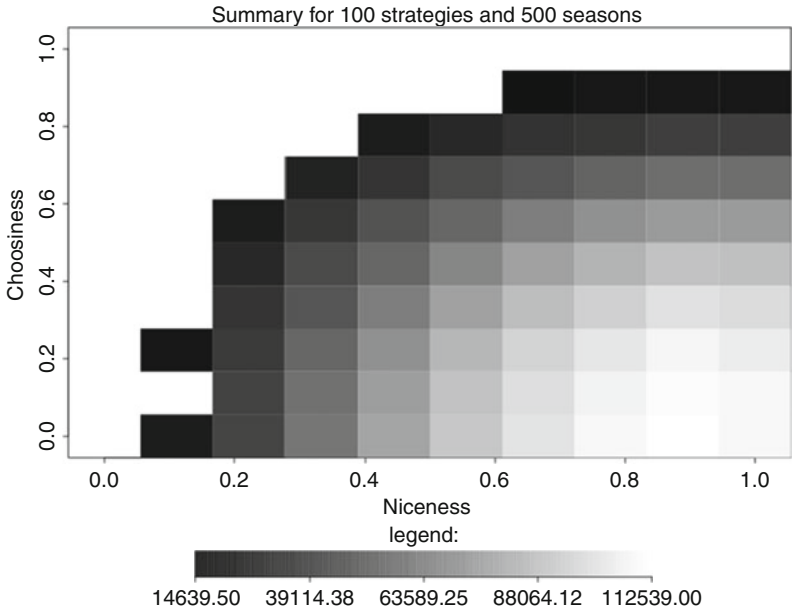

Figure 3. One hundred strategies after 500 seasons and .5 meeting cost. The *white spaces* indicate the dead strategies.

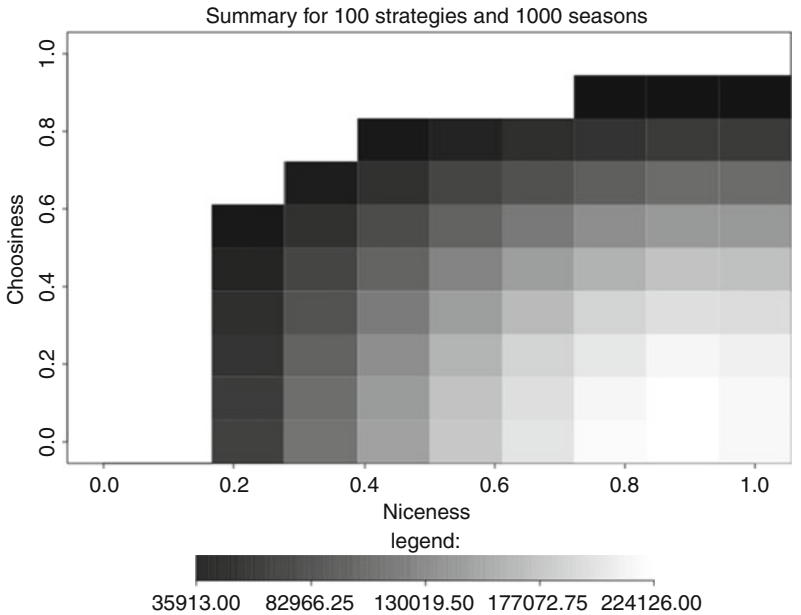

Figure 4. One hundred strategies after 1,000 seasons and .5 meeting cost. The nicest strategies continue to perform well.

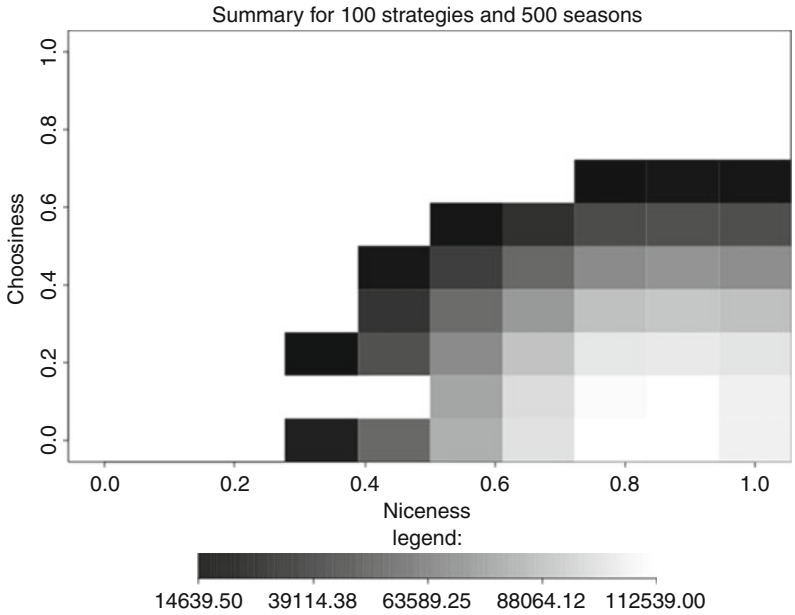

Figure 5. One hundred strategies after 500 seasons and 1.0 meeting cost. The nicest strategies are not doing quite as well.

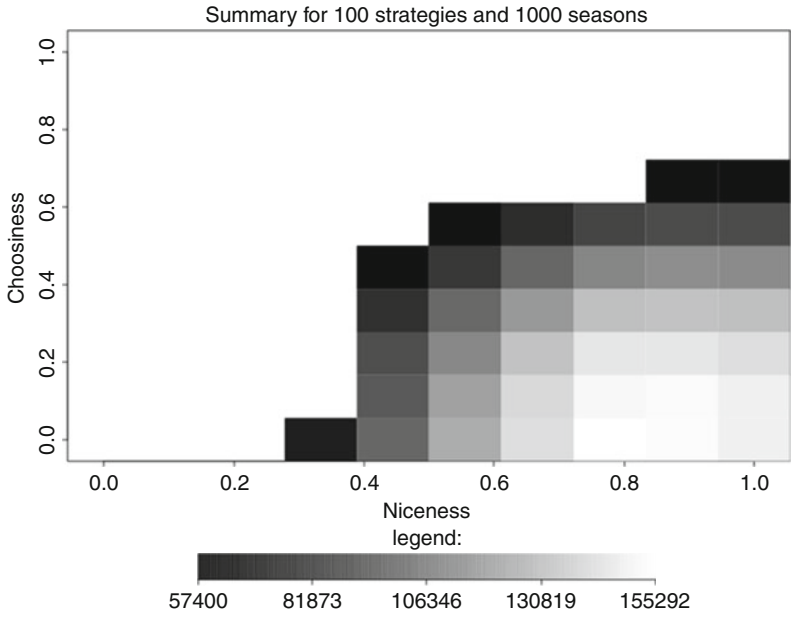

Figure 6. One hundred strategies after 1,000 seasons and 1.0 meeting cost. The surviving strategies have stabilized to compose a group that is fairly nice and not choosy.

nicer and more indifferent, the leaders become further removed from the (1, 0) standard. This is evidence for thinking that the nicest and least choosy strategies do very well (relative to the group) only when there are plenty of quite nasty and choosy strategies in the group. Another way of stating this phenomena is to note that as the meeting cost rises, the average niceness and the leaders' niceness both move toward convergence, while the average choosiness and the leaders' choosiness similarly converges. Under conditions of significant environing strain, very high niceness is not rewarded as much. Again, this result seems to also correspond to the way that cooperativeness diminishes somewhat when life is very hard.

The fourth phenomenon is observed in tournaments with both modest meeting costs and permitted births (available space regrettably forbids more graphs). The group gradually evolves toward an overall average higher niceness and an average lower choosiness. The most striking long-term stable trait of the group is how most of the strategies cluster to converge around a fairly high niceness and a fairly low choosiness.

The IRPG game exemplifies one way to show how the members of group who can monitor reputations and engage in voluntary cooperations will display, all other things being equal, an evolutionary convergence upon a relatively high niceness and a relatively low choosiness. This convergence on reciprocal cooperativeness remains quite robust even under conditions of moderate strain, and extreme strain only diminishes cooperativeness to a degree without disappearing. These features are the marks of a society well on its way to conducting itself morally as we have defined morality in previous sections, since moral conduct (more so than just etiquette) expects a habitually high degree of cooperativeness and a habitually high willingness to cooperate with most if not all other members of the group.

These remarkable traits by themselves are not sufficient for judging that these strategies are fully moral in the sense specified in earlier sections. Some features of moral practices are not modeled in this version of the IRPD game, such as imitative and instructive education and more exacting punishments than just shunning. However, these limitations are consistent with the way that this IRPD game can attempt to model hominid conduct. Furthermore, adding education and punishment to the IRPD game should produce similar overall results displaying the advantages of nice reciprocal cooperation, since imitating better strategies and supplementing punishment tend to enhance widespread cooperation (see Boyd and Richerson, 1992; Henrich and Boyd, 2001; Fehr and Fischbacher, 2004; Alexander, 2007; Marlowe, 2009). The role of punishment, from simple shunning and shaming to direct infliction of physical harm, is as crucial to indirect reciprocity as it is to morality and law, as would be evolutionarily expected. Sripada (2005) claims that only punishments and not reciprocity's benefits are behind morality, but this cannot be right, since agents would not long suffer the costs of enforcing something that itself offers few benefits.

Third-party punishment is probably deeply connected to the emergence of in-group morality. Third-party punishment would be a conspicuous display of fidelity to the group's welfare, but it would presumably require group sanction

(include group pity for the victim but not the transgressor) to produce net benefits to punishers (Okimoto and Wenzel, 2011). Furthermore, precisely because the retaliation by the punished back against punishers can diminish punishment's effectiveness (and hence effective reduce the situation to a zero-sum game of whom can dominate the other – see Janssen and Bushman, 2008), genuine morality would depend on a high degree of internalization so that just punishment is usually passively accepted. This dependency on internalization may account for the intuitive moral rightness of retribution. In small societies, the desire to avoid shaming and suffering from group sanction would have to powerfully instilled and maintained. In large societies, the development of law to reinforce important moral norms is commensurate with the emergence of police to wield sufficient force to forestall retaliation by those unrestrained by internalized morals.

7. Natural Morality

Early in hominid evolution, shunning and shaming would have been the primary means of punishing enforcement, along with intermittent physical violence. Our preliminary analysis of a basic IRPD game indicates that shunning would be very effective by itself, and other indirect reciprocity studies show that physical punishment with group sanction would only heighten enforcement and obedience. It is therefore reasonable to find both protomorality and the emergence of universal moral habits in those groups able to engage in non-zero-sum patterns of cooperation and to track each others' performance with just a small amount of information about reputation. The IRPD game successfully models something of what early hominid groups were likely capable of doing. The dominance and stability of such habitual social niceness does make a good practical fit with the sort of forms of intense and widespread cooperation our hominid ancestors were developing.

Despite intriguing modeling of the development of moral cooperation, we cannot forget how these capacities for reciprocal cooperation and protomorality would still be "in-group" features emerging among members of a social group. Even as cruelty or betrayal are subsiding within hominid tribes over hundreds of thousands of years as social intelligence increases to decrease reputation errors, those nasty deeds can robustly survive between tribes. Familiarity and closeness remain essential to our evolved sense of morality, as they have remained essential to charity. Efforts to expand the range of morality would hence require further reductions to errors of social judgment, expansions of who shall count as part of the "in-group," and enhancements of the sense of closeness to others through such things as new technologies of communication. And the field of ethics has indeed typically focused of these factors; cognitive psychology has also recommended increasing the availability of reliable information about other people, their reputations, and their social interactions (Pollock and Dugatkin, 1992; Paolucci and Conte, 2009).

Furthermore, once humans were able to cognitively appreciate how the norms of moral cooperation were modifiable, they could take some control over their habits and try to deliberate enhance morality's effectiveness at making cooperation even more reliably beneficial for all. Such experimental efforts must have been halting and unsteady, yet not without practical value, since humans did not abandon the effort and social morality eventually came to dominate human life. Without having to suddenly invent compassion, niceness, fairness, cooperativeness, civility, and trust, humans did intensify efforts to deliberately instill strong respect for morality's norms through operant conditionings and educational training, so that everyone would be more likely to habitually and voluntarily comply in all situations even if motivational feelings, strategic benefits, or punishments are not there.

Moral cooperation appears to be so consistently of higher benefit to all members of a social group in the long run that no one would be smarter for reverting to nothing but ruthless zero-sum games or outright harmful treachery. Of course, morality is hardly the only kind of practice that would enjoy long-term survival; we are just the sort of species in which multiple "strategies" would be distributed across a population. The habits of morality, installed by instinct and instilled by instruction, will always be statistically distributed: people will occasionally be mean, short-sighted, and selfish. Morality is designed to prevent, so far as possible, lapses by individuals into hurtful conduct and unintelligent noncooperation and to foster firm reforms of nonconforming members. In any society, of course, actual conduct will only approximately track that society's moral expectations, but those noticeable deviances are the exceptions that prove the normative rule.

Morality appears to be just the kind of low-information and largely habitual practice that could stably emerge among agents capable of recognizing many individuals, observing others' interactions, and tracking their reputations. Social intelligence leads to moral intelligence, and in return, morality improves society. In essence, during moral cooperation's long unconscious evolution in social groups, it represented long-term wisdom for most members even if individuals could not yet cognitively appreciate that fact. Furthermore, if indirect reciprocity morality can be propelled by group selection over long stretches of time (as suggested by Boyd and Richerson, 1990; Soltis et al., 1995; Nowak, 2011), in-group moral cooperation can also serve as a smart survival strategy for competing groups of hominids. The fact that no human society in recorded history has entirely abandoned morality of its own accord further indicates morality's durable value. Morality among hominids displayed a functional design by natural evolution, and our human moralities now display the imprint of our own redesigns as well.

Moral naturalism appears to have the resources needed to account for the origin and gradual development of the human practice of morality, so that no great leaps in emotional, cognitive, or spiritual abilities need to be postulated. Moral naturalism also appears to be able to account for why humans living in large societies put so much deliberative effort in ethical redesigns of moralities

and their enforcement with law. During the past 10,000 years, dramatically larger societies have been suffering from the inherited limitations of morality and getting obsessed with ethics and law, precisely because morality is naturally so emotional, limited to familiar in-groups, controlled by perceived reputation, and yet so essential to the needed expansion of nonviolent encounters and reciprocal cooperations among strangers.

8. Acknowledgments

Jeffrey Miecznikowski's helpful expertise in biostatistics at the Roswell Park Cancer Institute is responsible for the current coding for implementing the IRPD game and the graphs displaying its results. This chapter is also deeply indebted to Richard Carrier for his insightful criticisms and suggestions which resulted in many improvements and to Liz Stillwaggon Swan who patiently encouraged every stage of this chapter's progress from mere notion to fulfillment.

9. References

Alexander R (1987) The biology of moral systems. Aldine de Gruyter, New York
Alexander JM (2007) The structural evolution of morality. Cambridge University Press, Cambridge
Axelrod R (1984) The evolution of cooperation. Basic Books, New York
Barash D (2003) The survival game: how game theory explains the biology of cooperation and competition. Henry Holt, New York
Bekoff M, Pierce J (2009) Wild justice: the moral lives of animals. University of Chicago Press, Chicago
Boyd R, Richerson PJ (1989) The evolution of indirect reciprocity. Social Netw 11:213–236
Boyd R, Richerson P (1990) Group selection among alternative evolutionary stable strategies. J Theor Biol 145:331–342
Boyd R, Richerson P (1992) Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethol Sociobiol 13:171–195
Boyd R, Richerson P (2005) The origin and evolution of cultures. Oxford University Press, New York
Brinkmann S (2011) Psychology as a moral science: perspectives on normativity. Springer, New York
Carpenter J, Matthews P, Ong'Ong'a O (2004) Why punish? Social reciprocity and the enforcement of prosocial norms. J Evol Econ 14:407–429
Casebeer WD (2003) Natural ethical facts: evolution, connectionism, and moral cognition. MIT Press, Cambridge, MA
Colman A (1995) Game theory and its applications in the biological and social sciences, 2nd edn. Butterworth-Heinemann, Oxford
Darwall S, Railton P, Gibbard A (eds) (1997) Moral discourse and practice: some philosophical approaches. Oxford University Press, New York
de Waal F (1996) Good natured: the origins of right and wrong in primates and other animals. Harvard University Press, Cambridge, MA
de Waal F (2006) Morally evolved: primate social instincts, human morality, and the rise and fall of 'Veneer Theory'. In: Macedo S, Ober J (eds) Primates and philosophers: how morality evolved. Princeton University Press, Princeton, pp 1–80
de Waal F (2009) The age of empathy: nature's lessons for a kinder society. Harmony Books, New York
Doris JM et al (eds) (2010) The moral psychology handbook. Oxford University Press, Oxford

Fehr E, Fischbacher U (2004) Third party punishment and social norms. Evol Hum Behav 25:63–87
Flanagan O, Sarkissian H, Wong D (2008) Naturalizing ethics. In: Sinnott-Armstrong W (ed) Moral psychology, vol 1: the evolution of morality. MIT Press, Cambridge, MA, pp 1–26
Foot P (2001) Natural goodness. Oxford University Press, Oxford
Gazzaniga M (2005) The ethical brain: the science of our moral dilemmas. Harper Perennial, New York
Gigerenzer G, Selten R (eds) (2001) Bounded rationality: the adaptive toolbox. MIT Press, Cambridge, MA
Greene J (2008) The secret joke of Kant's soul. In: Sinnott-Armstrong W (ed) Moral psychology, vol 3: the neuroscience of morality. MIT Press, Cambridge, MA, pp 35–79
Haidt J (2003) The moral emotions. In: Davidson R et al (eds) Handbook of affective sciences. Oxford University Press, Oxford
Harris S (2010) The moral landscape: how science can determine human values. Free Press, New York
Hauert C, Traulsen A, Nowak M, Sigmund K (2008) Public goods with punishment and abstaining in finite and infinite populations. Biol Theor 3(2):114–122
Henrich J, Boyd R (2001) Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. J Theor Biol 208:79–89
Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R (2001) Cooperation, reciprocity and punishment in fifteen small-scale societies. Am Econ Rev 91:73–78
Hirshleifer J, Martinez Coll JC (1988) What strategies can support the evolutionary emergence of cooperation? J Conflict Resol 32:367–398
Joyce R (2001) The myth of morality. Cambridge University Press, Cambridge
Lillehammer H (2003) Debunking morality: evolutionary naturalism and moral error theory. Biol Philos 18:566–581
Marlowe FW (2009) Hadza cooperation second-party punishment, yes; third-party punishment, no. Hum Nat 20:417–430
Milinski M, Semmann D, Krambeck HJ (2002) Reputation helps solve the 'Tragedy of the Commons'. Nature 415:424–426
Nichols S (2004) Sentimental rules: on the natural foundations of moral judgment. Oxford University Press, Oxford
Nowak M (2006) Five rules for the evolution of cooperation. Science 314:1560–1563
Nowak M (2011) SuperCooperators: altruism, evolution, and why we need each other to succeed. Simon & Schuster, New York
Nowak M, Sigmund K (1998) The dynamics of indirect reciprocity. J Theor Biol 194:561–574
Nowak M, Sigmund K (2005) Evolution of indirect cooperators. Nature 437:1291–1298
Ohtsuki H, Iwasa Y (2006) The leading eight: social norms that can maintain cooperation by indirect reciprocity. J Theor Biol 239:435–444
Okimoto T, Wenzel M (2011) Third-party punishment and symbolic intragroup status. J Exp Social Psychol 47:709–718
Panchanathan K, Boyd R (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. Nature 432:499–502
Paolucci M, Conte R (2009) Reputation: social transmission for partner selection. In: Trajkovski G, Collins SG (eds) Handbook of research on agent-based societies: social and cultural interactions. Information Science Reference, Hershey
Pollock GB, Dugatkin LA (1992) Reciprocity and the evolution of reputation. J Theor Biol 159:25–37
Sachs JL, Mueller UG, Wilcox TP, Bull JJ (2004) The evolution of cooperation. Q Rev Biol 79:135–160
Sober E, Wilson DS (1998) Unto others: the evolution and psychology of unselfish behavior. Harvard University Press, Cambridge, MA
Soltis J, Boyd R, Richerson P (1995) Can group-functional behaviors evolve by cultural group selection—an empirical test. Curr Anthropol 36:473–494
Sripada CS (2005) Punishment and the strategic structure of moral systems. Biol Philos 20:767–789
Suzuki S, Akiyama E (2007) Evolution of Indirect Reciprocity in groups of various sizes and comparison with direct reciprocity. J Theor Biol 245:539–552
Taylor C, Nowak M (2007) Transforming the dilemma. Evolution 61:2281–2292

Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57
Trivers RL (1985) Social evolution. Benjamin Cummings, Menlo Park
Tullberg J (2004) On indirect reciprocity: the distinction between reciprocity and altruism, and a comment on suicide terrorism. Am J Econ Sociol 63:1193–1212
Wong D (2006) Natural moralities. Oxford University Press, New York
Wright R (2000) Nonzero: the logic of human destiny. Pantheon Books, New York