In The Routledge Handbook of Moral Epistemology, Aaron Zimmerman, Karen Jones, and Mark Timmons, editors. New York: Routledge Publishing.

9

THE EVOLUTION OF MORAL COGNITION

Leda Cosmides, Ricardo Andrés Guzmán, and John Tooby

1. Introduction

Moral concepts, judgments, sentiments, and emotions pervade human social life. We consider certain actions *obligatory*, *permitted*, or *forbidden*, recognize when someone is *entitled* to a resource, and evaluate character using morally tinged concepts such as *cheater*, *free rider*, *cooperative*, and *trustworthy*. Attitudes, actions, laws, and institutions can strike us as *fair*, *unjust*, *praiseworthy*, or *punishable*: moral judgments. Morally relevant sentiments color our experiences—*empathy* for another's pain, *sympathy* for their loss, *disgust* at their transgressions—and our decisions are influenced by feelings of *loyalty*, *altruism*, *warmth*, and *compassion*. Full-blown moral emotions organize our reactions—*anger* toward displays of disrespect, *guilt* over harming those we care about, *gratitude* for those who sacrifice on our behalf, *outrage* at those who harm others with impunity. A newly reinvigorated field, moral psychology, is investigating the genesis and content of these concepts, judgments, sentiments, and emotions.

This handbook reflects the field's intellectual diversity: Moral psychology has attracted psychologists (cognitive, social, developmental), philosophers, neuroscientists, evolutionary biologists, primatologists, economists, sociologists, anthropologists, and political scientists. Issues fundamental to each researcher's home field animate their questions. Investigators who started in philosophy might design experiments inspired by Kant, Mill, and Bentham to see when our moral judgments reflect deontic intuitions or deliberative reasoning about utilitarian consequences. Economists assume that decision-makers maximize their utility when making choices; when subjects in their experiments behave altruistically or punish free riders, they write utility functions that include "social preferences" to explain these choices. Evolutionary biologists model natural selection to understand which kinds of altruism it can favor. Anthropologists ask whether the content of morality varies capriciously across cultures or displays predictable patterns. Sociologists and political scientists see how trust and cooperation shape institutions and are, in turn, shaped by them. Developmentalists want to know whether infants have moral intuitions or begin life without them. Primatologists look for traces of human moral sentiments in our primate cousins, to ascertain the phylogeny of morality. Social and cognitive psychologists argue about the respective roles played by emotion and reasoning in moral judgment. Cognitive neuroscientists address the

emotion/reasoning debate by seeing which parts of the brain are activated when people make moral judgments. Neurologists ask whether moral judgment changes when people suffer damage to neural circuits that underwrite empathy. All interesting questions.

Here we illustrate how issues relevant to moral epistemology are studied in evolutionary psychology. As in the rest of the cognitive sciences, research in evolutionary psychology tests hypotheses about the architecture of the human mind: the information-processing systems that reliably develop in all neurotypical members of our species. It departs from traditional approaches by making use of an often overlooked fact: These cognitive systems evolved to solve problems of survival and reproduction faced by our hunter-gatherer ancestors. Theories of adaptive function, which specify these problems and what counts as a solution, are used to generate testable hypotheses about the design of these mechanisms. This research method has led to the discovery of many new, previously unknown features of attention, memory, reasoning, learning, emotion, decision making, and choice (e.g., Buss, 2015; Cosmides & Tooby, 2013; Lewis et al., 2017). And it has uncovered evidence of computational systems that are functionally specialized for regulating social interactions. Embedded in these evolved systems mechanisms of inference, judgment, and choice that generate intuitions about how we ought to treat others and how others ought to treat us: moral intuitions. That makes research on their design of direct relevance to moral psychology.

We are not claiming that all the intuitions, inferences, concepts, emotions, and judgments commonly thought of as "moral" are generated by one "moral module"—that is, by a single faculty of moral cognition that applies the same ethical principles to every domain of social life. The evidence accumulated so far—from evolutionary game theory, human behavioral ecology, paleoanthropology, studies of modern hunter-gatherers, and detailed research on cognitive processes—converges on a different view: What Darwin called the human moral sense arises from a number of different computational systems, each specialized for a different domain of social interaction. A single faculty of moral cognition is unlikely to exist because a single faculty of social cognition is unlikely to exist.

2. Why Would Selection Favor Multiple Systems Regulating Social Behavior?

Is all social behavior generated by a single cognitive system, a "faculty of social cognition"? The hypothesis that natural selection produced one system to handle functions as diverse as courting mates, helping kin, trading favors, and battling enemies is unlikely, for reasons we explain in this chapter. Ironically, a shorthand for talking about evolution and social behavior has contributed to the single faculty view.

In summarizing an evolutionary perspective, people occasionally say that organisms are "motivated to spread their genes." This creates the false impression that organisms have a single motivation—to spread their genes—and a general cognitive system that figures out how to do this. The same impression—that the mind is a blank slate equipped with a single goal—is created when animals are described as "choosing" to behave in ways that are adaptive—that is, in ways that increase the number of offspring that they (and their close relatives) eventually raise to reproductive maturity.

The mind does not—and cannot—work that way. It is impossible for a general purpose cognitive system—one devoid of programs specialized for different social domains—to

compute which course of action available to you now will maximize the number of offspring you (or your relatives) produce in the distant future. The full argument is beyond the scope of this chapter, but can be found in Cosmides and Tooby (1987, 1994) and Tooby and Cosmides (1990a, 1992). Organisms are not "motivated to spread their genes"—although it may sometimes appear that way.

It sows error and confusion to say (for example) that human mothers love and care for their children because they have a "selfish desire to spread their genes"—especially when discussing topics relevant to morality, such as altruism and selfishness. Maternal care does not exist in many species, but it does in primates: Primate mothers monitor their juvenile offspring, stay close to them, groom them, risk their own safety to protect them, and expend energy to feed them. Let's call the cognitive system that motivates this suite of behaviors *maternal love*. The care this system generated had consequences for a female primate's infants: It increased the probability that her offspring survived to reproductive age. Maternal love exists in our species because ancestral mothers who had this motivational system *had more surviving children* than those that did not, and those children inherited their mothers' adaptations for maternal care. Over deep time in the hominin line, motivational systems causing maternal care replaced alternative designs that led to neglect. We are descended from ancestral mothers who reliably developed adaptations that caused them to love, rather than neglect, their children. To say mothers love their children because they "want to spread their genes" posits an intention that does not exist and confuses levels of causation. Evolutionary biologists always distinguish adaptations—which are properties of phenotypes—from the selection pressures that caused them to evolve.

Distinguishing Proximate and Ultimate Causes

An organism's behavior is generated by *cognitive adaptations*: computational systems that were built by natural selection. The function of these evolved systems is to acquire information and use it to regulate behavior. Identifying these mechanisms and the information to which they are responding provides a causal explanation of the organism's behavior in the here and now (what biologists call a proximate explanation). But the computational properties of these adaptations exist as a downstream consequence of the manner in which they regulated behavior *in past environments*. Identifying the selection pressures that shaped these properties over deep time, and why they engineered a computational system with that design rather than an alternative design, provides a causal explanation, too: an ultimate (or functional) explanation.

The behavior produced by a mechanism has reproductive consequences: An animal with that mechanism might evade more predators, more accurately remember the location of fruiting trees, or choose more helpful cooperative partners than animals with a slightly different mechanism. Mutations can change the design of a mechanism, making it different from those found in other members of the species.[1] In a population of sexually reproducing organisms, a design feature that promotes reproduction better than existing alternatives leaves more replicas of itself in the next generation; over many generations, its relative frequency in the population increases until (usually) it replaces the alternative design (see below). For this reason, evolutionary biologists expect animal behavior to be regulated by computational systems that "tracked fitness" ancestrally—systems equipped

with features that produced adaptive (reproduction-promoting) behavior in the environments that selected for their design.

Ancestral Domains of Social Interaction

With this in mind, let us now return to the original question. Would selection have favored a single faculty of social cognition over alternative designs that existed ancestrally? Would a single faculty have replaced—and subsumed the functions of—a set of functionally distinct cognitive adaptations, each specialized for regulating behavior in a different domain of social interaction? To address this question, we first need to consider what kinds of social interactions our ancestors routinely engaged in.

The hunter-gatherer ancestors from whom we are descended engaged in many different types of social interaction. They hunted cooperatively, pooled risk by sharing food, formed long-term mating relationships, had short-term sexual liaisons, raised children, helped close kin, exchanged goods and favors, supported friends in disputes, competed for status, engaged in warfare, and weathered natural disasters together. Task analyses based on evolutionary game theory, human behavioral ecology, and what is known about ancestral environments indicate that what counted as adaptive (reproduction-promoting) behavior differed across these domains of social interaction and varied with the type of relationship (e.g., kin, mate, friend, rival).

- ~~For example, when~~ foraging success is determined more by luck than by effort, pooling risk by sharing food widely in the band benefits the individuals involved (Kaplan et al., 2012). Forming a risk pool is not adaptive, however, when productivity is a function of effort rather than luck. Evolved intuitions about when one "ought" to share, how much, and with whom can be expected to differ accordingly.
- Inflicting harm can promote the reproduction of individuals and their family members when the target is a man from a rival group, but it is rarely adaptive when he is a bandmate (Boehm, 2001; Wrangham & Peterson, 1997). The ethnographic record suggests that moral sentiments track this difference: Killing outgroup rivals commonly elicits pride and praise (Chagnon, 1992; Macfarlan et al., 2014); killing an ingroup member commonly elicits shame, anger, and censure (Boehm, 2012).
- Group cooperation unravels if free riders are not punished (Fehr & Gachter, 2000; Krasnow et al., 2015; Masclet et al., 2003; Yamagishi, 1986). But cooperation between two individuals can be sustained without punishing cheaters, when the option to switch partners exists (André & Baumard, 2011; Debove et al., 2015).
- Fidelity requires different actions (or inaction) depending on whether one is courting a mate or a political ally (Buss et al., 1992; Tooby & Cosmides, 2010).
- Reciprocating favors is necessary to maintain cooperation between friends (Trivers, 1971), but close relatives need not reciprocate help to continue receiving it (Hamilton, 1964).

These are just a few examples in which selection pressures differ radically across domains of social interaction. Each implies different inferences about how others "ought" to be treated and how others "ought" to treat us. This means that an evolved system designed to produce adaptive social inferences in one of these ancestral domains would fail to produce adaptive

inferences in the other domains. To produce adaptive behavior across *all* of these ancestral domains, each domain would have to activate a different set of cognitive adaptations.

The brain can be viewed as a set of evolved programs: computational systems that analyze situations and generate choices. Natural selection will not favor a single, cognitive system regulating choices—moral or otherwise—when programs tailored for tracking fitness in one domain (e.g., cooperative hunting, followed by sharing) require features that fail to do so in others (e.g., courtship, with competition for exclusive access to mates). To generate choices that tracked fitness ancestrally, the human cognitive architecture would need to have a number of different cognitive systems regulating social behavior, each tailored for a different class of social interactions (Bugental, 2000; Cosmides & Tooby, 1987, 1992, 1994; Haidt, 2012).

Multiple Systems to Implement Multiple Functions

Because what counts as the (adaptively) wrong thing to do differed from domain to domain, it is reasonable to predict the evolution of multiple systems regulating social interaction. Indeed, there should be as many domain-specific cognitive adaptations as there were ancestral domains in which the definitions of (evolutionarily) successful behavioral outcomes are incommensurate (for argument, see Tooby et al., 2005).

Because each of these systems evolved to regulate a different class of social interactions, each can be expected to have a different computational design—a different set of interlocking features, including domain-specialized concepts, inferences, motivational states, emotions, sentiments, and decision rules. When activated, these features should operate in concert, producing social intuitions—inferences, judgments, and choices—that would have promoted reproduction in the ancestral social contexts that selected for their design. The content of these social intuitions should vary across domains, however, depending on which adaptive specialization is activated. That will depend on cues in an individual's environment.

To be activated under the right circumstances, each domain-specialized system needs a front end designed to detect its target domain—a situation detector. Selection should favor situation detectors that use cues that were statistically associated with the target domain ancestrally. These cues can be very concrete (like the cry of a hungry infant, which triggers the flow of breast milk in a nursing mother) or quite abstract (like a string of foraging failures so long that it is unlikely to reflect bad luck). The perception that negative outcomes are due to bad luck should activate different sharing rules than the perception that these same failures are due to lack of effort on the part of those asking to share, for example (see below.) If we have cognitive adaptations with this design, then motivations to share—including intuitions about which distributions are "fair"—will shift in an orderly way with perceptions of luck versus effort.

3. Multiple Evolved Systems and Moral Pluralism

The search for a single overarching moral principle or value is appealing, whether it is a principle of utility or Kant's categorical imperative in its various formulations. But can a monist normative theory capture the complexity of human moral life? If social cognition is generated by multiple evolved systems, each with a different functional design, then it is unlikely that our moral intuitions can be systematized by a single principle or value.

Ideal utilitarianism and Kantian deontology were never advanced as descriptive theories of the mind, of course. But they have been proposed as guides to judgment and choice that humans *should* and therefore *can* use.

Practically speaking, moral principles have to escape from philosophy into the larger community to improve the moral quality of human life. Studies of cultural transmission show that ideas that engage evolved inference systems spread more easily from mind to mind than ones that do not. Boyer's (2001) analysis of which religious ideas become widespread, recurring across cultures and time, and which die on the vine illustrates this: Ideas that fail to engage our evolved intuitions fail to spread. If they survive at all, they become the esoterica of small communities of priests, monks, imams, rabbis, and other religious specialists. Esoteric debates among philosophers may give rise to moral rules and laws derived from a single, general moral principle, but these are unlikely to engage our evolved moral intuitions—they are more likely to collide with them instead (Cosmides & Tooby, 2008a, 2008b). That would limit their influence.

We can, of course, cognitively reframe situations to activate alternative evolved systems in an effort to live up to the ideals articulated by a general moral principle. If that is the goal, the descriptive theories of moral cognition emerging from evolutionary psychology suggest which cues and frames will be most effective.

But it may be easier for people to adopt and apply normative ideals and guides like those advanced by ethical intuitionists and moral sentimentalists, especially those who embrace pluralism (e.g., Audi, 2005; Gill & Nichols, 2008; Huemer, 2005; Ross, 1930). After all, a mind equipped with a set of cue-activated, domain-specialized systems regulating social interaction will generate moral inferences, judgments, sentiments, and intuitions that vary across social domains—creating pluralism of values and principles. These responses will also differ across time, situations, people, and cultures: Situation detectors respond to perceptions of local cues and facts, and these perceptions may differ depending on many factors, such as an individual's past experiences, knowledge, access to family, sociocultural environment—even that individual's current physiological state (e.g., hungry vs. sated—low blood glucose increases support for redistribution; Aarøe & Petersen, 2013). Moral intuitions will, therefore, vary accordingly.

Some argue that variation in "commonsense convictions"—moral diversity—undercuts the normative proposals advanced by ethical intuitionists (e.g., Singer, 2005; Greene, 2008). That argument does not hold, however, if the variation is systematic. Whale fins and chimp arms look different but, when seen in the light of evolution, the homology of bone structure is clear; Earth and Neptune have different orbits, but both are explained by Newton's universal law of gravitation. Diversity in the natural world resolves into patterns when the right conceptual framework is found. Moral diversity may also resolve into patterns when the architecture of our evolved computational systems is discovered, especially when this knowledge becomes integrated into theories of culture, institutions, and society (for examples, see Baumard & Boyer, 2013; Bloch & Sperber, 2002; Boyer, 2001, 2018; Boyer & Petersen, 2011; Cosmides & Tooby, 2006; Fiske, 1991; Henrich et al., 2012; Rai & Fiske, 2011).

That an adaptation evolved because it produced a particular (fitness-enhancing) pattern of behavior does not make that behavior moral—obviously. But the kind of species we are is surely relevant to ethical questions, if only because "ought" (arguably) implies "can." There is no point in arguing for the adoption of an ethical code if it violates evolved moral intuitions so profoundly that most humans will reject it.

For example, can human parents stop favoring their children over the children of strangers, as the most radical utilitarians say we must? And what would happen if they did? Let us assume for a moment that education, indoctrination, mindful meditation, or other cognitive technologies allow some parents to achieve true impartiality. What would this departure from an ancestrally typical social environment do to their children—mammals who evolved to expect a mother's love, whose social and emotional development depends on signals that their parents value them more than strangers? Would the children suffer emotional pain with each impartial act? Would they develop attachment disorders, turning into adults who cannot form long-term bonds or sustain a family life? No one knows for sure, but these outcomes are not implausible given clinical research on social development (e.g., Goldberg et al., 2000).

In the end, moral philosophers, politicians, and activists who argue in favor of particular rules, codes, and laws will have to decide what implications, if any, knowledge about human cognitive adaptations has for normative ethics, moral epistemology, and public policy. Our goal here is to explain some of the relevant selection pressures and point to research on the design of the mind that these theories of adaptive function have inspired.

4. Theories of Adaptive Function as Tools for Discovery

The lungs, the heart, the kidneys—every organ in the body has an evolved function, an adaptive problem it was designed[2] by natural selection to solve. Natural selection is a causal process that retains and discards features from an organism's design on the basis of how well they solve adaptive problems: cross-generationally enduring conditions that create reproductive opportunities or obstacles, such as the presence of predators, the need to share food, or the vulnerability of infants. Adaptive problems can be thought of as reproductive opportunities or obstacles in the following sense: *If* the organism had a property that interacted with these conditions in just the right way, *then* this property would have consequences that promote its reproduction relative to alternative properties. Over the long run, down chains of descent, natural selection creates suites of features that are functional in a specific sense: The elements are well-organized to cause their own reproduction in the environment in which the species evolved.

A correct theory of an organ's function explains its architecture down to the smallest detail and stimulates the discovery of new, previously unknown, features of its design. The lungs evolved for gas exchange, not (as previously thought) for cooling organs or mixing blood. This function explains the gross anatomy of the lungs (e.g., their similarity to bellows), identifies which features are byproducts (e.g., right and left sides have different shapes to accommodate the heart and liver, not for gas exchange per se), and generated hypotheses that led to the discovery of key functional properties. By searching for machinery well designed for solving problems of gas exchange, scientists found how the thinness and composition of alveolar membranes create a blood-air barrier, for example, and uncovered a computational system that regulates the rate and depth of breathing in response to changes in the partial pressure of O_2 and CO_2—information it extracts from arterial blood. These are *design features*, that is, properties selected for because they were well-engineered for solving that adaptive problem.

The brain is also an organ. Its function is not gas exchange, detoxifying poisons, or breaking down sugars; the brain is composed of neurons arranged into circuits because these circuits perform computations. The brain is composed of information-processing

devices—programs—that extract information from the environment and use it to regulate behavior and physiology. The question is, what programs are to be found in this organ of computation? What are the reliably developing, species-typical programs that reliably develop in most members of our species?

Theories of adaptive function are tools for discovering what programs exist and how they work. Each feature of each program that evolved to regulate behavior exists because the computations it generated promoted the survival and reproduction of our ancestors better than alternative computational features that arose during human evolutionary history. Natural selection is a hill-climbing process: over time, it assembles computational systems that solve problems that affected reproduction well, given the information available in the environments that selected for their design.

For more than 99% of our species' evolutionary history, our ancestors were foragers who made their living by gathering and hunting. To survive and reproduce, our ancestors had to solve many different, complex, adaptive problems, such as finding mates, protecting children, foraging efficiently, understanding speech, spotting predators, navigating, regulating body temperature, and attracting good cooperative partners.[3] Moreover, these problems had to be solved using only information that was available in ancestral environments.

Knowing this allows one to approach the study of the mind like an engineer. One starts by using theories about selection pressures and knowledge of ancestral environments to identify—and do a task analysis of—an adaptive information-processing problem. The task analysis reveals properties a program would have to have in order to solve that problem well; this suggests testable hypotheses about the design of programs that evolved to solve that problem. As in the rest of psychology, evolutionary psychologists conduct empirical research to find out whether systems with these computational properties exist in the brains of contemporary humans.

Moral psychology can be illuminated by research guided by theories of adaptive function. To illustrate this approach, we present one case in detail, followed by a cook's tour of research on cognitive adaptations for cooperation. The detailed case starts with the reproductive risks and opportunities that emerge for a species in which individuals interact frequently with their siblings.

5. Kin: Duties of Beneficence and Sexual Prohibitions

Clams never know their siblings. Their parents release millions of gametes into the sea, most of which are eaten. Only a few survive to adulthood, and these siblings are so dispersed that they are unlikely to ever meet, let alone interact. The ecology of many species causes siblings to disperse so widely that they never interact as adults, and siblings in species lacking parental care typically do not associate as juveniles either. Humans, however, lie at the opposite end of this spectrum. Hunter-gatherer children typically grow up in families with parents and siblings and live in bands that often include grandparents, uncles, aunts, and cousins. The uncles, aunts, and cousins are there because human siblings also associate as adults—like most people in traditional societies, adult hunter-gatherers are motivated to live with relatives nearby, if that is an option. Indeed, the hunter-gatherers from whom we are descended lived in small, semi-nomadic bands of 25–200 men, women, and children, most of them close relatives, extended family, and friends (Kelly, 1995).

That close genetic relatives frequently interacted ancestrally is an important fact about our species. Some of the best established models in evolutionary biology show that genetic relatedness is an important factor in the social evolution of such species (Hamilton, 1964; Williams & Williams, 1957). Genetic relatedness refers to the increased probability, compared to the population average, that two individuals will both carry the same randomly sampled gene, given information about common ancestors. The relatedness between two individuals (*i* and *j*) is typically expressed as a probability, r_{ij}, called the *degree of relatedness*. For humans, this probability usually has an upper bound around ½ (for full siblings; for parent and offspring) and a lower bound of zero (with nonrelatives).

The adaptive problems that arise for species who live with close genetic relatives are nonintuitive, biologically real, and have large fitness consequences. The most important ones involve mating and providing help.

6. Degree of Relatedness and Inbreeding Depression: Selection Pressures

Animals are highly organized systems (hence "organisms"), whose functioning can easily be disordered by random changes. Mutations are random events, and they occur every generation. Many of them disrupt the functioning of our tightly engineered regulatory systems. A single mutation can, for example, prevent a gene from being transcribed (or from producing the right protein). Given that our chromosomes come in pairs (one from each parent), a mutation like this need not be a problem for the individual it appears in. If it is found on only one chromosome of the pair and is recessive, the other chromosome will produce the right protein and the individual may be healthy. But if the same mutation is found on both chromosomes, the necessary protein will not be produced by either. The inability of an organism to produce one of its proteins can impair its development or prove fatal.

Such genes, called "deleterious recessives," are not rare. They accumulate in populations precisely because they are not harmful when heterozygous—that is, when they are matched with an undamaged allele. Their harmful effects are expressed, however, when they are homozygous—that is, when the same impaired gene is supplied from both parents. Each human carries a large number of deleterious recessives, most of them unexpressed. When expressed, they range in harmfulness from mild impairment to lethality. A "lethal equivalent" is a set of genes whose aggregate effects, when homozygous, completely prevent the reproduction of the individual they are in (as when they kill the bearer before reproductive age). It is estimated that each of us has at least one to two lethal equivalents worth of deleterious recessives (Bittles & Neel, 1994; Charlesworth & Charlesworth, 1999). However, because mutations are random, the deleterious recessives found in one person are usually different from those found in another.

These facts become socially important when natural selection evaluates the fitness consequences of mating with a nonrelative versus mating with a close genetic relative (for example, a parent or sibling). When humans reproduce, each parent places half of its genes into a gamete, which then meet and fuse to form the offspring. For parents who are genetically unrelated, the rate at which harmful recessives placed in the two gametes are likely to match and be expressed is a function of their frequency in the population. If (as is common) the frequency in the population of a given recessive is 1/1,000, then the frequency with which it will meet itself (be homozygous) in an offspring is only 1 in 1,000,000. In contrast,

if the two parents are close genetic relatives, then the rate at which deleterious recessives are rendered homozygous is far higher. The degree of relatedness between full siblings, and between parents and offspring, is ½. Therefore, each of the deleterious recessives one sibling inherited from her parents has a 50% chance of being in her brother. Each sibling has a further 50% chance of placing any given gene into a gamete, which means that for any given deleterious recessive found in one sibling, there is a 1/8 chance that a brother and sister will pass two copies to their joint offspring (a ½ chance both siblings have it times a ½ chance the sister places it in the egg times a ½ chance the brother places it in the sperm). Therefore, incest between full siblings renders one-eighth of the loci homozygous in the resulting offspring, leading to a fitness reduction of 25% in a species carrying two lethal equivalents (two lethal equivalents per individual × 1/8 expression in the offspring = 25%). This is a large selection pressure—the equivalent of killing one quarter of one's children. Because inbreeding makes children more similar to their parents, it also defeats the primary function of sexual reproduction, which is to produce genetic diversity that protects offspring against pathogens that have adapted to the parents' phenotype (Tooby, 1982).

The decline in the fitness of offspring (in their viability and consequent reproductive rate) that results from matings between close genetic relatives is called *inbreeding depression*. Although incest is rare, there are studies of children produced by inbreeding versus outbreeding that allow researchers to estimate the magnitude of inbreeding depression in humans. For example, Seemanova (1971) was able to compare children fathered by first-degree relatives (brothers and fathers) to children of the same women who were fathered by unrelated men. The rate of death, severe mental handicap, and congenital disorders was 54% in the children of first-degree relatives, compared to 8.7% in the children born of nonincestuous matings (see also Adams & Neel, 1967).

Both selection pressures—deleterious recessives and pathogen-driven selection for genetic diversity—have the same reproductive consequence: Individuals who avoid mating with close relatives will leave more descendants than those whose mating decisions are unaffected by relatedness. Thus natural selection will favor mutations that introduce motivational design features that cost-effectively reduce the probability of incest. In some primate species, this problem is solved by one sex (often males) leaving the natal group to join another troop. But for species like ours, in which close genetic relatives who are reproductively mature are commonly exposed to each other, an effective way of reducing incest is to make cues of genetic relatedness reduce sexual attraction. Incest is a major fitness error, and so the prospect of sex with a sibling or parent should elicit sexual disgust or revulsion—an avoidance motivation.

7. Kin Selection and Altruism

The theory of natural selection follows from replicator dynamics. Genes are a mechanism by which phenotypic features replicate themselves from parent to offspring. They can be thought of as particles of design: elements that can be transmitted from parent to offspring, and that, together with an environment, cause the organism to develop some design features and not others. Because design features are embodied in individual organisms, they can propagate themselves by solving problems that increase their bearer's reproductive success (very roughly, the number of offspring that reach reproductive age produced by that

individual). In evolutionary models, costs and benefits are usually reckoned as the average effects of a design feature on an individual's reproductive success. One way an organism can increase its reproductive success is by investing resources (e.g., metabolic energy, time) in ways that are likely to (i) produce more offspring in the future or (ii) improve the chances that existing offspring survive. The distinction between existing and future offspring does not matter for this analysis, so let's create a unit—offspring equivalents—for discussing the effects of a design feature on an individual's reproductive success.

A gene, however, can cause its own spread in two ways. It can produce a design feature that increases the reproductive success of (i) the individual it is in or (ii) other individuals who are more likely to carry that same gene than a random member of the population—that is, close genetic relatives. That probability is given by *r*, the degree of relatedness. This insight has implications for the evolution of social behavior, which were formalized in W. D. Hamilton's (1964) theory of kin selection.

When kin live in close association with one another, there are many opportunities for individuals to help their kin—to give them food, alert them to dangers, protect them from aggression, tend their wounds, lend them tools, argue in support of their interests, and so on. Given these opportunities, an organism can invest a unit of its limited resources in ways that increase its own reproductive success or that of its genetic relatives. The decision to allocate a unit of resource to a relative instead of one's own offspring has two net effects: It increases the relative's reproductive success (by an amount, B_{kin}, measured in offspring equivalents), and it prevents the helper from increasing its own reproductive success (an opportunity cost, C_{self}, representing offspring equivalents forgone). Consider, then, the fate of three alternative designs for a motivational system regulating decisions to help kin.

An individual with Design #1 invests all its resources in producing offspring of its own. When helping a genetic relative would decrease that individual's own reproductive success—that is, when $C_{self} > 0$—individuals with Design #1 decide to not help. Now imagine a population of individuals equipped with this design, living in an environment with a biologically plausible distribution of opportunities to help (the costs of providing help range from low to high, relative to the resulting benefits). In this population, a mutation emerges that causes the development of a different design. This new design motivates an individual to divide its resources between producing offspring of its own and helping its kin produce offspring. Under what conditions will this mutation spread?[4]

Consider first a mutation that produces Design #2, a motivational system that generates the decision to help kin whenever $B_{kin} > C_{self}$. Acts of help with these reproductive consequences increase the number of offspring produced by the kin member who received help, but that kin member may not have inherited the mutation that produced this design. For example, the probability that a full sibling inherited the same mutation—the one that produces Design #2—is only ½. When an individual with Design #2 allocates a resource to siblings, half of them do not have the mutation that produces this design; those that lack the mutation cannot pass it on to their offspring. This has consequences for the number of *copies of the mutation* in the next generation. When an individual with Design #2 gives a resource to its sibling, the average increase in new copies of the mutation produced *through the sibling* who received helped will be $\frac{1}{2}(0 + B_{sib})$. But the number of new copies produced *through the individual* who provided that help will be lower than if the individual had kept the resource—a decrease of C_{self}. (Technically the number of copies would be these values

(C_{self} and $½B_{sib}$) multiplied by the ½ chance a parent passes any given gene to its offspring, but this can be ignored because it is true for all parents—self and sibling both).[5]

For opportunities to help a sibling in which $B_{sib} > C_{self} > ½(B_{sib})$, individuals with Design #2 will decide to help their sibling. This decision allocates their resources in a way that causes a net decrease in the number of *copies of that design* in the next generation. Because siblings are only half as likely to have that mutation as the helper (self), the increase in copies of the mutation that result from this decision will not be large enough to offset the decrease in copies that would have been produced if self had kept the resource: C_{self} is greater than $½(B_{sib})$.[6] Given the same situation, individuals with Design #1 will not help their sibling: They will invest the resource in producing offspring of their own, who are twice as likely to have the gene for Design #1 as offspring produced by their sibling. When facing opportunities in this range, Design #1 produces more copies of itself than Design #2 does.

However, when facing opportunities to help where $½(B_{sib}) > C_{self} > 0$, Design #2 produces more copies of itself than Design #1 does. Individuals with Design # 1 allocate a resource to their own reproduction whenever $C_{self} > 0$, *no matter how large* B_{sib} *is*—that is, no matter how many more offspring their sibling would produce by using that resource. They make no tradeoffs between their own reproductive success and that of their siblings.

To see the consequences, let us consider situations in which keeping a unit of a resource allows an individual to produce one more offspring equivalent but giving it to a sibling would increase the sibling's reproductive success by three offspring equivalents. This is a situation in which $½(B_{sib}) > C_{self} > 0$. Given payoffs in this range, individuals with Design #1 keep the resource, whereas individuals with Design #2 invest it in their sibling, who has a ½ chance of carrying the mutation for Design #2. In the next generation, there will be $(3 \times ½) = 1.5$ copies of the mutation producing Design #2 for every copy of the gene for Design #1. When facing opportunities in this range, Design #2 produces more copies of itself than Design #1 does.

There is, however, a design that has the advantages of Design #2 without its disadvantages. Consider a mutation producing a third design. Individuals with this design are motivated to divide resources between self and kin, but their decision system *discounts the reproductive benefit to the close relative by the probability that this relative has inherited the same mutation*—which is given by $r_{self, kin}$. Individuals with this design help kin—they allocate a resource to kin rather than themselves—when the reproductive consequences helping are such that $(r_{self, kin} \times B_{kin}) > C_{self}$. This inequality is known as Hamilton's rule.

As before, let's assume that keeping a unit of a resource allows an individual to produce one more offspring equivalent. For opportunities to help a sibling in which $½(B_{sib}) > C_{self} > 0$, individuals with Design #1 decide to invest in their own offspring—producing one offspring equivalent—but individuals with the Hamiltonian design decide to allocate that resource to their sibling, producing >2 offspring equivalents (the same decision made by individuals with Design #2, with the same consequences). That decision translates into >1 copy of the Hamiltonian mutation in the next generation for every copy of the alternative allele, the gene that produces Design #1. This is the same relative advantage that the mutation producing Design #2 has over the gene for Design #1.

But for opportunities to help a sibling in which $B_{sib} > C_{self} > ½(B_{sib})$, individuals with the Hamiltonian design make the same decision as individuals with Design #1—they invest in their own offspring, producing one offspring equivalent. By contrast, individuals with

Design #2 decide to allocate that unit of resource to their sibling, thereby decreasing the number of copies of the mutation for Design #2 in the next generation. Individuals with Design #2 produce ½ offspring equivalent for every one produced by individuals with the Hamiltonian mutation. For reproductive payoffs in this range, the Hamiltonian mutation does as well as Design #1, and both produce more copies of their respective designs than Design #2 does.

In a population composed of individuals with Design #1, Design #2, and the Hamiltonian design, the mutation producing the Hamiltonian design will eventually outcompete the other two designs. This mutation promotes its own reproduction better than the existing alternatives by causing individuals who have it to make efficient tradeoffs between increasing their own reproductive success and the reproductive success of kin (who have the same mutation with probability $r_{self,\,kin}$).[7] For situations in which $r_{self,\,kin}(B_{kin}) > C_{self} > 0$, the Hamiltonian mutation produces more copies of itself than Design #1 produces, and does no worse than Design #2. For situations in which $B_{kin} > C_{self} > r_{self,\,kin}(B_{kin})$, the Hamiltonian mutation produces more copies of itself than Design #2 produces, and does no worse than Design #1. For this reason, the relative frequency of a Hamiltonian mutation can be expected to increase in the population over many generations, until it replaces the other designs: It will become universal[8] and species-typical.[9]

What about inflicting harm on kin when doing so would increase your own reproductive success? Small group living creates opportunities to benefit yourself at your sibling's expense—taking food from your sibling, seducing your sibling's mate, enhancing your reputation for formidability by publicly defeating your sibling in a fight, spreading gossip that makes you look like the better cooperative partner, and so on. The same Hamiltonian reasoning applies to motivations for not inflicting harm on kin when that would benefit oneself. A mutation for restraint will spread if it causes an organism to refrain from harming kin when the reproductive consequences are such that $C_{kin} \times r_{self,\,kin} > B_{self}$. That is, the decrease in the reproductive success of the relative needs to be discounted by the probability that the relative inherited the same mutation for restraint.

Reciprocation is usually necessary to select for adaptations for delivering benefits to nonrelatives (see below). But reciprocation is not necessary for selection to favor adaptations for helping kin. The motivational system need not trigger the inference that the sibling helped is obligated to reciprocate. For situations in which the payoffs satisfy Hamilton's rule, individuals with adaptations shaped by kin selection will be motivated to help their kin without any conditions. Kin selection will favor adaptations that produce unconditional altruism toward kin: Close relatives will not need to reciprocate help to continue receiving it (Hamilton, 1964; Williams & Williams, 1957).

8. Estimating and Representing Benefits and Costs: A Computational Requirement across Social Domains

Hamilton's rule is not a computational mechanism; it is not an algorithm that makes decisions. It describes selection pressures that can be expected to shape cognitive adaptations operative when organisms *do* make decisions about how to divide resources between self and kin. What computational properties would adaptations like this require?

In Hamilton's rule, C_i and B_i refer to the actual effects of an action on the reproductive success of individual *i*. But at the time an organism makes a decision, it does not—and cannot—know how that decision will affect its reproductive success in the future. This is true for every choice an organism makes: which food to eat, which mate to pursue, whether to freeze or flee when seen by a predator, whether to keep a resource or donate it to kin—all of them.

Making tradeoffs between options requires computational machinery that estimates the value of one option relative to another. For example, foragers—people who hunt and gather for a living—search for and harvest some plants and prey, while ignoring others. Their choices are systematic: their decisions can be predicted by models that assume they are optimizing calories obtained for a given amount of effort (Smith & Winterhalder, 1992). Knowing four variables allows behavioral ecologists to explain 50% of the variance in their decisions about which resources to pursue. Two involve effort: search time (how long until first encounter with the resource) and handling time (the time from first encounter to when the resource is ready to eat). The other two involve nutritive value: the resource's caloric density (calories/unit volume; e.g., avocado > cucumber) and typical volume (size of an animal or a resource patch). The success of these models implies the existence of psychological mechanisms that estimate effort and caloric value, plus mechanisms that use these values in realizing an organism's decision as to which resources to pursue. To have produced adaptive behavior ancestrally, the values that these mechanisms compute would have to use information that reflected the average reproductive consequences of choices in our ancestral past. For example, our taste for fats and sugars evolved because these chemicals were correlated with the caloric value of food, and they were difficult to acquire ancestrally. Tastes for fats and sugars caused foraging decisions and food choices that were adaptive (reproduction-promoting) ancestrally.

These tastes guide our food choices now too: that is why ice cream—a food high in fats and sugars—"tastes good." But these preferences, which caused adaptive behavior in the past, may be maladaptive now—they can lead to diabetes and early death in advanced market economies where foods high in fat and sugar are not only abundant, but available with low search and handling time at supermarkets and fast food restaurants.

To avoid confusion, it is important to distinguish reproductive costs and benefits in the past—that is, selection pressures—from costs and benefits as computed by an organism's evolved computational systems. We don't like ice cream more than oat bran because this preference promotes reproduction in the present; we like it now because design features causing preferences for fats and sugars promoted reproduction in the past. Humans, like other organisms, have computational systems that evolved to assign value to options we face.

If the selection pressures described by Hamilton's rule designed cognitive systems for deciding how to divide resources between self and kin, these systems would require input from other mechanisms, which estimate the costs and benefits of actions to self and others in a way that reflected average reproductive consequences in our ancestral past. Indeed, every theory of the evolution of social behavior assumes that mechanisms of this kind exist. The values computed by these mechanisms serve as input to cognitive adaptations for making social decisions—especially ones that make decisions about how we ought to treat others and how others ought to treat us.

The taste for fats and sugars is just one component of one value-computing system, and a very specialized component at that. Notice that evolved systems for computing food value cannot assign a fixed value to specific foods: The value computed for a given food should be higher when my blood sugar is low than when it is high, for example. When my blood sugar is high, and there are cues that my sister is hungry, my value-computing systems should estimate that the benefit she will derive from the venison I have will be higher than the cost to me of giving it to her. Nor can there be a fixed value for eating over other activities because adaptive behavior requires complicated tradeoffs. As one example: there are value-computing systems in women that prioritize sex over eating on days when conception is most likely (more specifically, on days when estrogen is high and progesterone low) and eating over sex on the days before menstruation (when estrogen is low and progesterone high; Roney & Simmons, 2017). This is not because women have a "motivation to spread their genes." Sex is pleasurable—and libido fluctuates with these hormone profiles—because adaptations with these features promoted reproduction in ancestral environments.

The design of value-computing systems is relevant to utilitarian theories of ethics, which assume that people can estimate the consequences of actions for the welfare of self and others. Surprisingly little is known, however, about how human minds estimate benefits and costs or how these are represented within and across domains. Input to systems that evolved for estimating the marginal benefit of keeping an additional unit of a resource versus giving it to another person should include many factors: the type of resource (food, time, energy, social capital, actions that carry a risk of death), each individual's age (younger individuals have more of their reproductive career ahead of them), health, current reproductive state, current nutritional status, relationship (e.g., mate, child), resources already available to self and other, the size of a resource to be divided (what an economist might call income), and so on (e.g., Burnstein et al., 1994). Evolved systems should be able to calculate the costs and benefits of options presenting themselves on the fly, because these inputs are not fixed variables—they can change quickly.

Whether the factors that serve as input to these calculations are morally justifiable is a question for moral epistemologists. For our purposes, we will assume that such systems exist and that they were designed to track fitness in ancestral environments. When we are discussing the design of adaptations that make social decisions, "costs" and "benefits" refer to the *perceived* values of resources or actions, i.e., the values *as computed by the mind of the individual who is making a decision*—not to the effects of these resources or actions on the lifetime reproductive success of the decision maker, its siblings, or anyone else.

9. A Kin Detection System: Computational Requirements

These two adaptive problems—inbreeding avoidance and kin-directed altruism—both require that close kin are treated differently than unrelated individuals. That requires some means of distinguishing one's close genetic relatives from people who are related distantly or not at all. A task analysis of this adaptive problem led to testable predictions about the presence and properties of a kin detection system: a neurocomputational system that is well engineered (given the structure of ancestral environments) for computing which individuals in one's social environment are close genetic relatives (Lieberman et al., 2007).

For each familiar individual, *j*, the kin detection system should compute and update a continuous variable, the *kinship index*, KI_j. By hypothesis, KI_j is an internal regulatory variable whose magnitude reflects the kin detection system's pairwise estimate of the degree of relatedness between self and *j*. The kinship index should serve as input to at least two different motivational systems: one regulating feelings of sexual attraction and revulsion and another regulating altruistic impulses. When KI_{ij} is high, it should up-regulate motivations to provide aid to *j* and down-regulate sexual attraction by activating disgust at the prospect of sex with *j*.

Ancestrally Reliable Cues to Genetic Relatedness

Detecting genetic relatedness is a major adaptive problem but not an easy one to solve. Neither we nor our ancestors can see another person's DNA directly and compare it to our own, in order to determine genetic relatedness. Nor can the problem of detecting genetic relatives be solved by a domain-general learning mechanism that picks up local, transient cues to genetic relatedness: To identify which cues predict relatedness locally, the mechanism would need to already know the genetic relatedness of others—the very information it lacks and needs to find.[10] Instead, the kin detection system must contain within its evolved design a specification of the core cues that it will use to determine relatedness—cues picked out over evolutionary time by natural selection because they reliably tracked genetic relatedness in the ancestral social world. This requires *monitoring circuitry*, which is designed to register cues that are relevant in computing relatedness. It also requires a computational unit, a *kinship estimator*, whose procedures were tuned by a history of selection to take these registered inputs and transform them into a kinship index. So what cues does the monitoring circuitry register, and how does the kinship estimator transform these into a kinship index?

For our hunter-gatherer ancestors, a reliable cue to relatedness is provided by the close association between mother and infant that begins with birth and is maintained by maternal attachment. Maternal perinatal association (MPA) provides an effective psychophysical foundation for the mutual kin detection of mother and child. It also provides a foundation for sibling detection. Among our ancestors, when an individual observed an infant in an enduring caretaking association with the observer's mother, that infant was likely to be the observer's sibling. To use this high-quality information, the kin detection system would need a monitoring subsystem specialized for registering MPA.

Although MPA allows older siblings to detect younger siblings, it cannot be used by younger siblings because they do not exist at the time their older siblings were born and nursed. This implies that the kin detection system's psychophysical front end must monitor at least one additional cue to relatedness. The cumulative duration of coresidence between two children, summed over the full period of parental care until late adolescence, is a cue that could be used to predict genetic relatedness—an expansion and modification of an early ethological proposal about imprinting during early childhood (Shepher, 1983; Westermarck, 1891/1921; Wolfe, 1995).

Hunter-gatherer bands fission and fuse over time, as their members forage and visit other bands; this means individuals frequently spent short periods of time with unrelated or distantly related persons. However, hunter-gatherer parents (especially mothers) maintained

close association with their dependent children in order to care for them. Siblings, therefore, maintained a higher-than-average cumulative association with each other within the band structure. As association is summed over longer periods of time, it monotonically becomes an increasingly good cue to genetic relatedness. This invites the hypothesis that the kin detection system has a system for monitoring duration of coresidence between *i* and *j* during *i*'s childhood, and that its output is particularly important for younger siblings to detect older siblings.

10. Does a Kin Detection System Regulate Sibling Altruism and Sexual Aversion?

To compute the kinship index, the kin detection system requires: (1) monitoring circuitry designed to register cues to relatedness (MPA, coresidence during childhood, possibly other cues) and (2) a computational device, the kinship estimator, whose procedures have been tuned by a history of selection to take these registered inputs and transform them into a kinship index—the regulatory variable that evolved to track genetic relatedness.

If these cues are integrated into a single kinship index—that is, if the kinship index for each familiar individual is a real computational element of human psychology—then two distinct motivational systems should be regulated by the same pattern of input cues. For example, when *i* is younger than *j*, *i*'s kinship index toward *j* should be higher the longer they coresided during *i*'s childhood. As a result, *i*'s levels of altruism and sexual aversion toward *j* will be predicted by their duration of childhood coresidence.

Lieberman et al. (2007) tested these hypotheses about the computational architecture of human kin detection by quantitatively matching naturally generated individual variation in two predicted cues of genetic relatedness—maternal perinatal association and duration of coresidence during childhood—to individual variation in altruism directed toward a given sibling and opposition to incest with that sibling. When the MPA cue was absent (as it always is for younger siblings detecting older siblings), duration of childhood coresidence with a specific sibling predicted measures of altruism and sexual aversion toward that sibling, with similar effect sizes. When the MPA cue was present (which is possible only for older siblings detecting younger siblings), measures of altruism and sexual aversion toward the younger sibling were high, regardless of childhood coresidence.

The fact that two different motivational systems are regulated in parallel by the same cues to genetic relatedness implicates a single underlying computational variable—a kinship index—that is accessed by both motivational systems. Finally, the results imply that the architecture includes a kinship estimator, which integrates the cues to produce the kinship index. If the effects of the cues were additive, there could be a direct path from each cue to each motivational system. Instead, when both cues were available, the more reliable cue—maternal perinatal association—trumps coresidence duration. That these two cues interact in a non-compensatory way implies they are being integrated to form a variable, which then serves as input to the systems motivating altruism and sexual aversion. This pattern of cue activation has since been replicated six times with measures of altruism, in samples drawn from the US (California, Hawaii), Argentina, Belgium, and a traditional Carib society practicing horticulture (Sznycer et al., 2016); effects of coresidence duration on altruism and sexual aversion were also tested and confirmed among unrelated adults who had been

co-reared during childhood on a kibbutz in Israel, in communal children's houses where groups of similar-aged peers slept, ate, and bathed (Lieberman & Lobel, 2012).

This entire computational system appears to operate nonconsciously and independently of conscious beliefs. When beliefs about genetic relatedness conflict with the cues this system uses (as they do when people have coresided with stepsiblings or unrelated peers), the motivational outputs (caring, sexual disgust) are shaped by the cues, not the beliefs. Coresidence duration predicts sexual aversion and altruism toward stepsiblings (Lieberman et al., 2007) and toward genetically unrelated people raised together on kibbutzim in Israel (Lieberman & Lobel, 2012).

11. Moral Sentiments about Siblings

How the kinship index is computed creates systematic variation in the strength of moral intuitions across individuals. First, it regulates the strength of moral proscriptions against sibling incest. Second, it regulates how often people sacrifice to help their siblings and how willing they are to do so.

There is debate among evolutionary psychologists about why people are motivated to endorse moral prescriptions and prohibitions. The adaptive functions proposed include relationship regulation (Baumard et al., 2013; Cosmides & Tooby, 2006, 2008a; Fiske, 1991; Rai & Fiske, 2011), binding cooperative groups together via adherence to sacred values (Haidt, 2012), promoting within-group cooperative norms (Boehm, 2012; Boyd & Richerson, 2009), reducing the costs associated with taking sides in other people's disputes by coordinating condemnation with other third parties (DeScioli & Kurzban, 2013), creating a local moral consensus favorable to realizing the individual's preferences (Kurzban et al., 2010; Tooby & Cosmides, 2010), and mobilizing coalitions of individuals with similar interests to treat the enforcement of norms as a collective action (Tooby & Cosmides, 2010; see also Boyer, 2018).

Most of these theories converge in proposing a link between disgust and morality. The emotion of disgust is reliably elicited by cues correlated with the presence of pathogens (e.g., rotting corpses, vomit, mold, (someone else's) bodily fluids) and by the prospect of sex with genetic relatives and other partners whose value as a potential mate is low (for review, see Tyber et al., 2013). Its evolved function is to motivate one to avoid actions or objects that would have imposed fitness costs ancestrally (and now). Moral prohibitions specify actions and objects to be avoided. A default heuristic to moralize actions that are felt to be against one's interests would connect disgust to moral prohibitions, as would attempts to promote self-serving prohibitions by portraying actions as disgusting (Tooby & Cosmides, 2010; Tyber et al., 2013). Disgust and moral prohibitions both tag actions as wrong to do. Many empirical studies confirm this link: Actions that elicit disgust are often moralized, and actions that are judged morally wrong sometimes elicit disgust (for reviews, see Haidt, 2012; Lieberman & Patrick, 2018; Tyber et al., 2013).

Disgust, Morality, and Sibling Incest

Haidt and colleagues showed that the intuition that brother-sister incest is morally wrong is so strong that it persists even when the scenario described has removed all practical reasons for avoiding it (e.g., contraception was used, no one else will know, it was consensual; Haidt, 2012). This resistance to reasoning, which Haidt refers to as "moral dumbfounding," supports a

claim relevant to ethical intuitionists: that we directly apprehend (or seemingly apprehend) the wrongness of certain actions, in a process akin to a perceptual experience (Stratton-Lake, 2016).

The strength of these intuitions varies systematically, however, with factors that regulate the kinship index (Fessler & Navarrete, 2004; Lieberman et al., 2003, 2007). In the studies reviewed earlier, Lieberman et al. (2003, 2007) asked people to rank how morally wrong 19 acts were, where the list included consensual sex between siblings (third parties, not oneself). The pattern was the same as for disgust. When the MPA cue was absent, moral wrongness judgments tracked duration of coresidence with opposite-sex siblings; for subjects with younger opposite-sex siblings, they tracked the presence of the MPA cue. As the disgust-morality link predicts, this result is specific to coresidence with *opposite-sex* siblings: coresidence with same-sex siblings does not predict moral judgments about sibling incest at all. This result speaks against any counterexplanation that attributes harsher judgments to factors (such as having a traditional family structure) that are correlated with siblings having coresided for a long time (Lieberman et al., 2003).

What about people who are not biological siblings, yet raised together? Lieberman and Lobel (2012) had the same kibbutz-raised adults rate (1) disgust at the idea of sex with their opposite-sex peers, (2) how morally wrong it would be for kibbutz classmates to have sex, and (3) how morally wrong it would be for a brother and sister to have sex. Coresidence duration with opposite-sex *peers* did not predict judgments of how wrong *sibling* incest is. It predicted how morally wrong it would be for *kibbutz classmates* to have sex and how disgusting they would find sex with their opposite-sex peers. These disgust and wrongness ratings were strongly correlated, as expected. But a causal pathway from coresidence duration to disgust to morality was confirmed by mediation analyses. The correlation between coresidence duration and sexual disgust remained high when ratings of moral wrongness were controlled for statistically. But controlling for sexual disgust erased the link between coresidence duration and moral wrongness; sexual disgust fully mediated the relationship between the coresidence cue and moral wrongness judgments.

Some societies have explicit prohibitions (rules, norms, or laws) against incest with harsh punishments for transgressions, whereas other societies either lack explicit prohibitions or, if these exist, lack harsh punishments. Why does this cross-cultural variation exist, if an evolved mechanism causes most people to find the prospect of sex with siblings distasteful? In a comprehensive review of the ethnographic literature, Fox (1965/1984) showed that explicit prohibitions against incest are most common in societies where the sexes are segregated during childhood—a practice that results in opposite-sex siblings spending a lot of time apart. Explicit prohibitions are either absent, or accompanied by a relaxed attitude, in societies where opposite-sex siblings live in close association during childhood.

Research on the computational architecture of the kin detection system demonstrates that there is variation in moral intuitions about sex with siblings (and peers!). But this variation is systematic: when analyzed in light of this evolved system, the moral diversity resolves into patterns.

Altruism and Duties of Beneficence toward Siblings

The ethnographic record supports the prediction that kin selection will create adaptations motivating unconditional altruism toward kin (Fiske, 1991). Altruism toward kin is

widespread across cultures, and so is the ethic that kin ought to treat each other with generosity "without putting a price on what they give" and without demanding "strictly equivalent returns of one another" (Fortes, 1970, 237–238; see Fiske, 1991). Like the wrongness of incest, this obligation is directly apprehended (or seemingly apprehended): "kinship is felt to be inescapable, presupposed, and unproblematic. . . [it] inherently involves a fundamental moral and affective premise of amity, solidarity, concern, trust, prescriptive altruism manifested in generosity, loving, and freely sharing" (Fiske, 1991, 354).

Notice, however, that the architecture of the kin detection system creates systematic variation in altruism toward siblings *within* a culture. The same cues that regulate moral intuitions about incest—MPA and coresidence duration—regulate how often people sacrifice to help their siblings (as measured by favors done in the last month) and their willingness to incur large costs (such as donating a kidney), whether they are true biological siblings, stepsiblings, or unrelated children being raised together on a kibbutz.

But will moral intuitions about how much you should sacrifice to help a sibling be the same *within* a family? No. Trivers' (1974) application of kin selection theory to family relationships predicts that different family members will have different intuitions about how much you *should* sacrifice to help your sibling—that is, different views about *your* duties of beneficence. Trivers' insight was that kin selection will favor adaptations for social negotiation within the family. If so, then adaptations in each family member should be designed to weight their estimates of the costs to you and the benefits to your sibling by that family member's kinship index toward each of you. For ease of exposition, we will assume that each family member has a kin detection system that computed a kinship index that reflects $r_{self,j}$, the degree of relatedness between that individual and family member *j*.

Let's say you could take a costly action that benefits your sister, a full sibling. Let's also assume that you, your sister, and your mother all agree on the magnitude of C_{you} and B_{sister} that will result from your helping (each of you has an evolved program that evaluates this action prospectively, generating estimates of these values; see §8, "Estimating and Representing Benefits and Costs"). All else equal, your adaptations will motivate you to help your sister when $C_{you} < ½ \times B_{sister}$. Because adaptations in your sister will also have been shaped by kin selection, they will discount C_{you} by her kinship index, which reflects her degree of relatedness to you; the intuition produced by her adaptations will be that you *ought* to help her when $½C_{you} < B_{sister}$ (but not when $½C_{you} > B_{sister}$; kin selection implies there will be limits on the costs she is willing to impose on you). The magnitude of your mother's kinship index will be the same for both of you—her degree of relatedness to each of you is ½. So her kin-selected adaptations will generate the intuition that you should help your sister whenever $½B_{sister} > ½C_{you}$; that is, your mother will encourage you to help when $B_{sister} > C_{you}$. Her opinion will be shared by every other member of the family, *k*, for whom $r_{k,you} = r_{k,sister}$: your father, your other full siblings, your grandparents, and their children (your uncles and aunts).

These kin-selected adaptations can be expected to include moral concepts, such as "ought" and "should": the feelings and opinions they generate are about how you *ought* to treat your sister, how she *deserves* to be treated by you. When you act otherwise—in reality or when contemplating options prospectively—these adaptations can be expected to activate moral emotions. These emotions themselves have evolved functions, which are reflected in their computational design (Tooby & Cosmides, 2008; Tooby et al., 2008).

Research on the computational architecture of anger and shame provide examples (e.g., Sell, Sznycer, Al-Shawaf, et al., 2017; Sell et al., 2009; Sznycer et al., 2016). Prospectively, the moral emotions produce evaluations of alterative actions, used in social decision making (e.g., Sznycer et al., 2016). After the fact, they recalibrate variables used by social decision-making systems (e.g., correcting estimates of another person's need—the relevant costs and benefits), and motivate relevant behaviors (such as bargaining for better treatment (Sell et al., 2017) or apologizing and withdrawing socially to avoid being further devalued by others (Sznycer et al., 2016)). For example, you may feel guilt at the prospect of helping too little or, after the fact, in response to information from your mother or sister that you underestimated your sister's need (you may experience regret when you realize you overestimated her need). Your sister and mother may grow angry when you help less than their adaptations calculated you should, motivating them to communicate this to you (argue), threaten to withdraw benefits from you, or otherwise incentivize you to treat your sister better in the future (Sell et al., 2017).

As a thought experiment, let's assume that taking the action under review benefits your sister by 5 notional units ($B_{sister} = 5$). As long as $C_{you} < 2.5$ (i.e., $< \frac{1}{2}B_{sister}$), you will be motivated to help, and your mother and sister will agree that you should. The same reasoning implies that the three of you will agree that you should *not* help your sister when $C_{you} > 10$ (siblings are selected to refrain from imposing too much harm on one another: for values over 10, $\frac{1}{2}C_{you} > 5 = B_{sister}$). There will be a consensus about whether you *ought* to help your sister—moral connotation intended—when the cost to you is smaller than $\frac{1}{2}B_{sister}$ and larger than $2B_{sister}$.

The three of you will have different opinions, however, when the cost to you is between these values (in this example, when $2.5 < C_{you} < 10$). Mom weighs your welfare and your sister's equally, but your sister discounts costs to you by ½; so when $5 < C_{you} < 10$, your sister will want your help, but your mother will think this is too much to ask, and you will too. But when $2.5 < C_{you} < 5$, your mother will have the intuition that you ought to help your sister (because $B_{sister} > C_{you}$) and your sister will agree, but you will feel they are expecting too much from you. That is, your three brains will generate conflicting intuitions about what you ought to do—what your duties of beneficence are—when $\frac{1}{2}B_{sister} < C_{you} < 2B_{sister}$. There is no single solution in this range that will generate moral agreement.

In fact, the same logic implies that your own moral intuitions will shift when the shoe is on the other foot—that is, when your sister has the option of helping you. The idea that she should help you will seem reasonable to you when $C_{sister} < 2B_{you}$, but she will disagree when $C_{sister} < \frac{1}{2}B_{you}$. You will seem like a hypocrite: You will expect more from her than you were willing to give her in an equivalent situation (Kurzban, 2012). When you are a mother with two daughters, your intuitions about their duties of beneficence toward one another will change again: When $C_i < B_j$, you will feel that daughter *i* should help daughter *j*. Your own moral intuitions should change because different cognitive adaptations will be activated depending on your role in this family drama: when you are the sibling doing the helping, when you are the sibling being helped, and when you are the mother of two children.

What does this mean about moral intuitions regarding duties of beneficence toward full siblings? Moral consensus within the family should emerge for actions involving a wide range of costs to you and benefits to your sibling—especially when $C_{you} < \frac{1}{2}B_{sib}$ (you should

help her) and when $C_{you} > 2B_{sib}$ (don't help her, it is too much to ask). For values in these ranges, the evolved psychologies of various family members can be expected to generate similar intuitions, feelings, or opinions about how you *ought* to treat your sibling—about what counts as the *right* thing to do. These mechanisms may also generate a moral consensus in a society—including all-things-considered judgments—when people contemplate situations of this kind prospectively, read literature, or try to decide who was wrong in a conflict between siblings.

When the costs and benefits of *i* helping sibling *j* fall between those values—that is, when everyone agrees that $½C_i > B_{sib} > 2C_i$—moral conflict is likely. Dissension will arise within the family about *i*'s duties toward sibling *j*, and your own intuitions will vary depending on whether you are the parent, the helper, or the sibling who can be helped. The evolved psychologies of various family members can be expected to generate different intuitions, feelings, or opinions about how you ought to treat your sibling for values in this range. An outcome that is morally satisfying to one sibling will feel unfair to the other.

In this analysis, there is no impartial point of view. What counted as a reproduction-promoting "strategy" differed for ancestral mothers, self, and siblings, and the psychologies designed by those selection pressures can be expected to vary accordingly. The evolved psychology of mothers weights each sibling equally, not because her age and experience have made her an impartial judge but because equal weightings promoted the reproduction of mothers in our ancestral past. Her preferences were shaped by the same causal processes that produced unequal weightings in self and sibling.

If the goal of ethical theorists is to create normative theories of broad applicability, analyses like this can help. Across a wide variety of situations, different family members can be expected to have very similar moral intuitions about an individual's duties of beneficence toward their siblings. The prospects for developing normative principles that capture the intuitions of everyone involved are promising for situations like these. But when actions are perceived as implying costs and benefits in the intermediate range just discussed, the moral intuitions of family members can be expected to differ, and there is no outcome that will feel fair to all involved. Although this variation is systematic, the search for a normative principle that captures the intuitions of everyone involved is likely to fail. Theorists could accept this, adopting a normative principle for such cases that captures some, but not all, moral intuitions. Or they could decide that these situations involve matters of personal taste—like preferences for chocolate versus peppermint ice cream—that fall outside the scope of moral judgment.

12. The Evolution of Cooperation: A Cook's Tour

Biologists have been interested in the "problem of altruism" since the 1960s (Williams, 1966). They define the problem thus: How can selection favor an adaptation that causes an organism to systematically behave in ways that decrease the organism's own reproductive success while therein increasing the reproductive success of another individual? Actions or structural features that have these consequences by design[11] are defined as "altruistic" in biology. Kin selection is one of several solutions to biology's "problem of altruism." But kin selection cannot explain the evolution of adaptations causing altruism toward individuals who are not kin.

Adaptations causing unconditional altruism toward kin have evolved in many taxa. Altruism toward individuals who are not kin is less common zoologically, and requires adaptations that are different from those that generate altruism toward kin. In most models, altruism must be conditional to evolve. Even group selection models require the effects of altruism to fall differentially on ingroup members (e.g., Bowles & Gintis, 2013; Boyd & Richerson, 2009; McElreath & Boyd, 2006). The extent to which humans in all cultures cooperate with individuals who are not genetic relatives is among our most zoologically unusual features. What follows is a brief tour of adaptations for conditional cooperation, including social exchange in its various forms: reciprocal altruism, cooperation for mutual benefit, trade (Axelrod & Hamilton, 1981; Barclay, 2013, 2016; Baumard et al., 2013; Fiske, 1991; Noë & Hammerstein, 1995; Trivers, 1971); cooperation in groups, especially collective action (Boyd & Richerson, 2009; Tooby et al., 2006); deep engagement (banker's paradox) relationships (Tooby & Cosmides, 1996); and risk-pooling cooperation, which can apply within cooperative dyads or groups (Kaplan & Hill, 1985; Kaplan et al., 2012). The adaptive problems that need to be solved for cooperation to evolve in these various forms are similar (but not identical); solving them requires computational systems with domain-specialized concepts, representational formats, reasoning systems, and moral sentiments. Many deontic concepts and implicit moral rules are embedded in these systems (Curry, 2015). They are also relevant to virtue ethics, providing a basis for understanding which kinds of characteristics are likely to be treated as virtues across cultures and time.

13. Evolutionary Game Theory and the Analysis of Social Behavior

Game theory is a tool for analyzing strategic social behavior—how agents might behave when they are interacting with others who can anticipate and respond to their behavior. Economists have used it to analyze how people respond to incentives present in a well-defined situation. These models typically assume rational actors who calculate the payoffs of alternative options (anticipating that other players will do likewise) and choose the option that maximizes the actor's own payoff (but see Hoffman et al., 1998).

The social behavior of other people was as relentless a selection pressure as predators and efficient foraging. To specify these selection pressures more precisely, evolutionary biologists adopted game theory as an analytic tool, too (Maynard Smith, 1982). Evolutionary game theory requires no assumptions about deductive reasoning or economic rationality; indeed, it can be usefully applied to cooperation among bacteria or fighting in spiders. It is used to model interactions among agents endowed with well-defined decisions rules that produce behavior that is contingent on features of the situation (especially the behavior of other agents). Although these decision rules are sometimes called "strategies" by evolutionary biologists, this is a term of art: no deliberation by bacteria (or humans) is implied (or ruled out) by this term. Whether the decision rules being analyzed are designed to regulate foraging, fighting, or cooperating, the immediate payoffs of these decisions, in food or resources, are translated by the modeler into the currency of offspring produced by the decision-making agent, and these offspring inherit their parents' decision rule. In evolutionary game theory, a decision rule or strategy that garners higher payoffs leaves more copies of itself in the next generation than alternatives that garner lower payoffs. By analyzing the reproductive consequences of alternative decision rules, evolutionary biologists can determine

which strategies natural selection is likely to favor and which are likely to be selected out (eliminated from the population).

14. The Evolution of Cooperation between Two Unrelated Individuals: Constraints from Game Theory

The evolution of adaptations for cooperation between unrelated individuals is tricky, even when only two individuals are involved and they can interact repeatedly. Yet social exchange—an umbrella term for two-party cooperation in its many forms—is a ubiquitous feature of every human society. In evolutionary game theory, it is often modeled as a repeated "prisoner's dilemma" (PD) game, with two agents who are not genetic relatives.[12] In each round of a PD game, an agent must decide whether to cooperate or defect—to provide a benefit of magnitude B to the other agent (at cost C to oneself) or refrain from doing so. In a PD game, $B - C > 0$ for both agents. In evolutionary game theory, the choice made by each agent is specified by a decision rule (a "strategy"), and different agents are equipped with different decision rules. When an agent reproduces, its offspring have the same design—the same decision rule—as the parent (with high probability; many models allow mutations, i.e., a small probability that an offspring has a different decision rule from its parent).

In a repeated PD, two agents play many rounds during a single generation. When it is time to reproduce, the benefits and costs each agent earned during these rounds—payoffs that can be thought of as calories acquired *vs*. expended, favors garnered *vs*. given, changes in status from winning *vs*. losing a fight—are translated into offspring in the next generation. The relative number of offspring each agent produces before it "dies" is proportional to the payoffs it earned during that generation: agents with designs that earned higher payoffs produce more offspring relative to agents with designs that earned lower payoffs.[13] That is, the agents' choices have consequences for their reproductive success (as in models of kin selection). This process is repeated for many generations, so the modeler can determine which decision rules—which strategies—increase in relative frequency and which are eliminated from the population.

Imagine a population of agents participating in a series of PD games. Each agent is equipped with one of two possible decision rules: *always cooperate* or *always defect*. *Always cooperate* causes unconditional cooperation: agents with this design incur cost C to provide their partner with benefit B, regardless of how their partner behaves in return. The other decision rule, *always defect*, accepts benefits from others but never provides them, so it never suffers cost C. When two unconditional cooperators interact, their payoff is positive, because $B - C > 0$. When two defectors interact, they get nothing—they are no better or worse off than if they had not interacted at all. But every time a cooperator interacts with a defector, the cooperator suffers a net loss of C (because it pays cost C with no compensating benefit) and the defector, who incurred no cost, earns B (the benefit provided by the cooperator).

Now imagine that these agents are randomly sorted into pairs for each new round and there are n rounds during a generation. Because assortment into pairs is random, the probability that an agent is paired with a cooperator is p, the proportion of cooperators in the population; the probability an agent is paired with a defector is $(1 - p)$, the proportion of defectors in the population. The *always defect* rule never suffers a cost, but it earns B every

time it is paired with an agent who always cooperates, which is $n \times p$ times; thus $np \times B$ is the total payoff earned by each defector that generation. In contrast, the *always cooperate* rule suffers cost C in every round, for a total cost of $n \times C$. It earns B only from the *np* rounds in which it meets another cooperator, for a total benefit of *np*B. Hence, $n(pB - C)$ is the total payoff earned by each cooperator that generation. These payoffs determine the relative number of offspring each agent produces in the next generation. Because offspring have the same design as their parents with high probability, these payoffs also determine the relative number of copies of a design in the next generation (mutations are random with respect to design). Because $npB > npB - nC$, the *always defect* design will leave more copies of itself in the next generation than the *always cooperate* design. As this continues over generations, unconditional cooperators will eventually disappear from the population, and only defectors will remain. In an environment where the only alternative is a design that always cooperates, *always defect* is an evolutionarily stable strategy (ESS), but *always cooperate* is not.

Although strategies that cause unconditional cooperation fail, models in evolutionary game theory show that decision rules that cause cooperation can evolve and be maintained in a population by natural selection if they implement a strategy for *conditional* cooperation—a strategy that not only recognizes and remembers (at least some of) its history of interaction with other agents, but uses that information to cooperate with other cooperators and defect on defectors. (One example of a strategy with these properties is *tit-for-tat*, a decision rule that induces cooperation on the first move, after which its adherent does whatever its partner did on the previous move; Axelrod & Hamilton, 1981; Axelrod, 1984.) Conditional cooperators remember acts of cooperation and cooperate in response, so they provide benefits to one another, earning a payoff of $(B - C)$ every time they interact. Because the cooperation of one elicits future cooperation from the other, these designs cooperate with one another repeatedly, and the positive payoffs they earn from these interactions accumulate over rounds. In this, they are like unconditional cooperators. The difference is that conditional cooperators limit their losses to defectors. The first time a conditional cooperator interacts with a particular defector, it suffers a one-time loss, C, and the defector earns a one-time benefit, B. But the next time these two individuals meet, the conditional cooperator defects and does not resume cooperation unless its partner responds by cooperating. As a result, designs that defect cannot continue to prosper at the expense of designs that cooperate conditionally. Designs that cooperate conditionally harvest gains in trade from interacting repeatedly with one another; interactions between designs that defect do not produce these gains in trade. Because reproduction of a design is proportional to the payoffs it earns, designs that induce conditional cooperation produce more copies of themselves in the next generation than designs that induce defection. It is a prediction of this approach that over many generations a population that begins with both designs will gradually replace designs that always (or usually) defect with designs that cooperate conditionally.

Defectors are often referred to as *cheaters* in two-party reciprocation or social exchange. The results of evolutionary game theory suggest that cognitive adaptations for participating in social exchange can be favored and maintained by natural selection, but only if they implement some form of conditional cooperation. To do so, they require design features that detect and respond to cheaters so defined.

Notice that evolutionary game theory analyzes which *designs* are favored by selection. In this context, "cheaters" are individuals who cheat in situations involving social exchange by virtue of their cognitive design: They are agents with decision rules that cause them to take benefits provided by another agent without providing what the other agent wanted. (For the purposes of this analysis, it does not matter whether the agent's failure to reciprocate was caused by an intentional choice or by the calibration of a nonconscious ("subpersonal") mechanism.) Not all failures to reciprocate indicate a cheater: conditional cooperators will sometimes fail to reciprocate because they suffered bad luck (e.g., their hunt failed; injury prevented them from foraging) or made a mistake. Models from evolutionary game theory show that withdrawing from cooperation with these individuals is an adaptive error (e.g., Panchanathan & Boyd, 2003; see below on generosity and free riders).

15. Detecting Cheaters and Reasoning about Social Exchange

Using constraints from game theory and knowledge about the behavioral ecology of hunter-gatherers, Cosmides and Tooby developed *social contract theory*: a task analysis specifying (1) the adaptive problems that arise in two-party reciprocation and (2) the properties a computational system would need to solve them (Cosmides, 1985; Cosmides & Tooby, 1989, 2008a). As in the discussion of adaptations for kin-directed altruism, we assume that the human cognitive architecture has adaptations for computing the value of resources, actions, and situations to self and other. In what follows, "benefit" and "cost" refer to these computed values—to mental representations generated by the reasoner. Research on reasoning about social exchange shows that procedures for detecting cheaters operate on abstract representations of costs and benefits (e.g., Cosmides & Tooby, 2008a, 2008c).[14]

The provision of benefits needs to be conditional for social exchange between unrelated individuals to evolve: you deliver a benefit to an agent *conditional* on that agent satisfying some requirement of yours (providing a direct benefit or creating conditions that benefit you). Whether the understanding is left implicit or agreed to explicitly, this contingency can be expressed as a *social contract*, a conditional rule that fits the following template: *If you accept benefit B from me, then you must satisfy my requirement R*. A cheater is an individual who has taken benefit *B* without satisfying requirement *R* and has done so by design, not by mistake or incapacity.

Understanding social exchange and detecting cheaters requires some form of conditional reasoning. The human cognitive architecture may well include subroutines that implement the inferences of first-order logic or a relatively domain-general deontic logic. But the inferences that these subroutines make will systematically fail to detect cheaters (for explanations, see Cosmides & Tooby, 2008a; Fiddick et al., 2000). Because these logics are content blind, they are insensitive to what *P* and *Q* refer to in "if *P* then *Q*." Reasoning adaptively about social exchange requires rules of inference that are specialized for this domain of social interaction. These inference procedures need to be content sensitive—they need to operate on representational primitives such as *benefit* to *agent 1*, *requirement* of *agent 2*, *obligation*,[15] *entitlement*, *intention to violate*, *perspective* of *agent i*. And they must include a subroutine that looks for *cheaters*, a specialized moral concept. The empirical evidence shows that situations of social exchange do, in fact, activate the very specialized representations, inference

rules, and violation detection procedures required to reason adaptively in this domain (for detailed reviews of the evidence, see Cosmides & Tooby, 2008a, 2008c, 2015). Reasoning about social exchange dissociates—both functionally and neurally—from reasoning about deontic rules so similar to social contracts that no other theory distinguishes between them.

The neurocognitive system activated by situations involving social exchange generates inferences about how people "ought" to treat one another in interactions of this kind and triggers negative evaluations of cheaters. For this reason, it can be thought of as a moral reasoning system: a very specialized one.

16. Specialized Inferences: An Example

The inferences of this specialized moral reasoning system—the *social contract algorithms*—diverge from the inferences of classical first order logics (Cosmides & Tooby, 2008a; Fiddick et al., 2000). In first order logic, "if *P* then *Q*" does not imply "if *Q* then *P*." But social contract algorithms license an inference like this when *P* and *Q* refer to the benefits and requirements of agents in social exchange. For example, when Ana says "Bea, if you babysit my son, I will give you a sack of avocados" and Bea agrees, that implies "If Bea accepts the avocados from Ana, she is obligated to babysit Ana's son." When Bea babysits, this triggers the inference that she is entitled to the avocados and Ana is obligated to provide them; when Ana provides the avocados to Bea, it triggers the inference that she is entitled to have Bea babysit and Bea is obligated to do so. These are moral inferences, which we spontaneously make in situations of social exchange.

A content-general deontic logic (a logic of obligation, entitlement, and prohibition) will not generate the adaptively correct pattern of inferences either. For situations involving social exchange, [1] "If you accept benefit *B* from agent X, then you are obligated to satisfy X's requirement," implies [2] "If you satisfy agent X's requirement, then you are entitled to the benefit *B* that X offered to provide" (and vice versa). But consider a slightly more general version of [2] that operates outside the domain of social exchange: "If you satisfy requirement R, then you are entitled to E." This cannot imply "If you get E, then you are obligated to satisfy requirement R" without violating our moral intuitions. For example, "If you are an American citizen, you have a right to a jury trial" does not imply "If you get a jury trial, then you are obligated to be an American citizen" (Cosmides & Tooby, 2008a).

17. Cheater Detection

In first order logic, a conditional of the form "if P then Q" is violated when P is true and Q is false, that is, by the co-occurrence of a true antecedent and a false consequent (P & not-Q). The classical rules of deductive inference (like modus ponens and modus tollens) operate on the antecedent and consequent of a conditional, no matter what they refer to; these procedures are blind to benefits, requirements, and agents with perspectives. Social contract algorithms employ a very specific concept of violation—cheating—that does not map onto the concept of violation used in first order logic. Cheating is taking the benefit from an agent without satisfying that agent's requirement. A cheater is someone who does this by design, not by mistake.

Consider again Ana's offer, "If you babysit my son, I will give you a sack of avocados." Which acts count as cheating depends on whether we adopt the perspective of Ana or

Bea. Ana cheated if she accepted Bea's help babysitting but did not give her avocados; Bea cheated if she accepted avocados from Ana but did not babysit. That's not because social contracts are biconditional: if Bea does not babysit, Ana has not cheated if she decides to give Bea avocados anyway, nor has Bea cheated if she babysits but then decides she doesn't need any more avocados.

Studies with the Wason selection task, a tool developed in cognitive psychology to study conditional reasoning, show that social contracts activate a cognitive mechanism that looks for cheaters, not for logical violations. In this task, subjects are presented with a conditional rule and then asked to look for cases that could violate it. If you ask whether Ana violated the rule, they investigate cases in which Bea babysat (P) and cases in which Ana did not give her avocados (not-Q). This response—P & not-Q—is the same response subjects would give if they were looking for logical violations, reasoning with rules of logic. But they are not. If you ask instead whether Bea violated the rule, they investigate cases in which Bea accepted avocados (Q) and cases in which she did not babysit (not-P). Q & not-P is logically incorrect, but it is the correct answer if you are looking to see if Bea cheated (Gigerenzer & Hug, 1992). The same pattern of adaptively correct but logically incorrect answers can be elicited by having Ana express her offer like this: "If I give you avocados, then you must babysit my son." Subjects asked whether Ana violated the rule still investigate occasions when Bea babysat (Q) and occasions when Ana gave her nothing (not-P). In formal logic, a true consequent (Q) with a false antecedent (not-P) does not violate a conditional rule, but these cases are the right ones to investigate if you want to know if Ana cheated (Cosmides, 1985, 1989).

The most telling results come from experiments that present the same social contract rule but vary information about those who are in a position to violate it (Cosmides et al., 2010). Social contract theory predicts an adaptive specialization that looks for cheaters, not innocent mistakes. Cues relevant to this distinction regulate when subjects detect cases that violate the rule. First, intentional violations activate cheater detection, but innocent mistakes do not. Second, violation detection is up-regulated when potential violators would get the benefit regulated by the rule and down-regulated when they would not. Third, cheater detection is down-regulated when the situation makes cheating difficult—when violations are unlikely, the search for them is unlikely to reveal those with a disposition to cheat. Parametric studies show that each of these cues independently contributes to violation detection (Cosmides et al., 2010; for discussion see Cosmides & Tooby, 2015).

This provides three converging lines of evidence that the mechanism activated by conditionals expressing a social contact is not designed to look for general rule violators, or deontic rule violators, or violators of social contracts, or even cases in which someone has been cheat*ed*. This mechanism does not look for violators of social exchange rules in cases of mistake—not even in cases when someone has accidentally benefited by violating a social contract. The mechanism activated has a narrow focus: It looks for violations of social contracts when this is likely to lead to detecting cheaters—defined as individuals who take a benefit that was conditionally offered by an agent while intentionally not meeting that agent's requirement. Its computational design fits the adaptive function—detecting designs that cheat—like a key fits a lock: It is a cheater detection mechanism.

18. Partner Choice versus Partner Control

Once you detect cheaters, then what? Broadly speaking, there are two possibilities: You can choose to cooperate with a different partner or you can try to reform the partner who cheated so she cooperates with you more in the future. Evolutionary biologists refer to these options as partner choice versus partner control (e.g., Barclay, 2013; Nöe & Hammerstein, 1994, 1995; Schino & Aureli, 2017). Switching partners can be the less costly strategy if alternative cooperative partners are available to you (assuming the costs of finding and establishing a new cooperative relationship are low). If they are not, the best strategy may be to reform the partner you have. The two main bargaining tools available to incentivize better behavior are to (1) inflict harm (i.e., punish) or (2) withdraw cooperation until the partner starts cooperating again—the tactic employed by tit-for-tat strategies (TFT). Both partner control tools are costly: the first risks injury, the second entails forgone opportunities to forge a profitable cooperative relationship with someone else. Threatening to use these tactics is less costly than deploying them, but when the partner herself can switch, it risks losing a partner who provides net benefits despite subtle cheating (under-reciprocating).

When the repeated prisoner's dilemma was first used to model the evolution of cooperation, partners were paired with one another randomly (Trivers, 1971; Axelrod & Hamilton, 1981). In that environment, partner choice is not an option. Strategies for conditional cooperation arising from these models employ partner control tactics (e.g., tit-for-tat). Lately, inspired by Nöe and Hammerstein's (1994, 1995) early papers distinguishing partner control from partner choice models of the evolution of cooperation, there has been a florescence of new research on cooperative partner choice in "biological markets." Both kinds of model have implications for moral psychology, and can shed light on morality-relevant puzzles arising from results in behavioral economics.

19. Puzzles from Behavioral Economics

Cooperation can be studied in the laboratory by having people interact in games in which the monetary payoffs for different choices are carefully controlled—dictator games, prisoner's dilemma style games, bargaining games (e.g., the ultimatum game), trust/investment games, public goods games, and others. When behavioral economists used these methods to test predictions of game theory, they found that people in small groups do not act as if they are maximizing immediate monetary payoffs (e.g., Hoffman et al., 1998; Smith, 2003). In a one-shot interaction with anonymous others, *Homo economicus* models predict no generosity, no cooperation, no trust, and no punishment. Yet people give more, cooperate more, trust more, and punish defections more than these models predict, even when the experimenter tells them that the interaction is one-shot and anonymous. Why? According to both economic and evolutionary game theory, repeated interactions are necessary for behaviors like this to evolve.

To some, this "excess altruism" is evidence that the psychology of cooperation was shaped by group selection rather than selection operating on individuals. According to these models, groups that included individuals with psychological designs that led them to suffer costs to punish defectors would maintain higher levels of within-group cooperation and, therefore, outcompete groups without such individuals. Although individuals with designs that punish defectors will have lower fitness than members of their group who cooperate without punishing, this "strong reciprocity" design spreads because groups with

these individuals replace groups that lack them (Bowles & Gintis, 2013; Boyd et al., 2003; Gintis, 2000; Gintis et al., 2003).

But are these behaviors really *excess* altruism—that is, beyond what can be explained by selection on individuals for direct reciprocity? Selection does not occur in a vacuum: The physical and social ecology of a species shape the design of its adaptations, and our hunter-gatherer ancestors lived in small, interdependent bands that had many encounters with individuals from neighboring bands. Adaptations for direct reciprocity evolved to regulate cooperation in an ancestral world in which most interactions were repeated. The high prior probability that any given interaction will be repeated should be reflected in their design. In fact, models of this social ecology show that meeting an individual once is a good cue that you will meet again (Krasnow et al., 2013).

This has been called the "Big Mistake" hypothesis by advocates of group selection—who characterize this position as saying that our adaptations are "mistaking" one-shot interactions for repeated ones (e.g., Henrich & Henrich, 2007, 91). Critics of the Big Mistake hypothesis argue that one-shot interactions were common enough in the lives of ancestral hunter-gatherers to select against cooperation in these situations. On this basis, they argue that the Big Mistake hypothesis is mistaken. But is it? Partner control and partner choice models provide evidence that the "Big Mistake mistake" is not a mistake.

20. Partner Control and the Evolution of Generosity

Agent-based simulations are widely used to study the evolution of cooperation by partner control. In most cases, the behavioral strategies are particulate—they do not have internal cognitive components that can evolve—and the simulation environment has either one-shot or repeated interactions, but not both. But what happens if these strategies have components that can evolve, and the social environment includes both one-shot *and* repeated interactions, as in real life? It turns out that generosity in one-shot interactions evolves easily when natural selection shapes decision systems for regulating two-person reciprocity (exchange) under conditions of uncertainty (Delton et al., 2011).

In real life, you never know with certainty that you will interact with a person once and only once (until the moment before you die). Categorizing an interaction as one-shot or repeated is always a judgment made under uncertainty, based on probabilistic cues (e.g., am I far from home? Does she speak with my accent? Did he marry into my band?). In deciding whether to initiate a cooperative relationship, a contingent cooperator must use these cues to make tradeoffs between two different kinds of errors: (i) false positives, in which a one-shot interaction is mistakenly categorized as a repeated interaction, and (ii) misses, in which a repeated interaction is mistakenly categorized as one-shot. A miss is a missed opportunity to harvest gains in trade from a long string of mutually beneficial interactions. In a population of contingent cooperators, the cost of a miss is usually much higher than the cost of a false positive.

To see this, consider agents who defect on a new partner when they believe the interaction is one-shot, but play TFT when they believe they will repeatedly interact with the new partner. Let's assume repeated interactions last for only five rounds on average, and a round produces very modest gains in trade: $b = 3$, $c = 1$, so the payoff for mutual cooperation is $(b - c) = 2$ and the cost of cooperating with a defector is $c = 1$. The cost of a false

positive error is $c = 1$: the payoff for an agent who cooperates, (wrongly) assuming this will be a repeated interaction, with a partner who defects. But notice that the cost of a miss is 10 times greater (5 rounds × $(b - c)$): When the agent defects, (wrongly) assuming it is a one-shot interaction, its new partner defects in return, inaugurating a chain of reciprocal defections. Even this is an underestimate: Given that humans have relationships that span decades, an average of five rounds for repeated interactions is low. When the average number of rounds for repeated interactions is 10 (still low), the opportunity cost of a miss is the failure to harvest a payoff of 20 ($10(b - c)$)—in this case, the cost of a miss is 20 times larger than the $c = 1$ cost of a false positive. When misses are more costly than false positives, it can be better to have fewer missed opportunities at the price of more false positives—cases in which agents cooperate in one-shot interactions.

Using agent-based simulations, Delton et al. (2011) show that under a wide range of conditions, individual level selection favors computational designs that decide to cooperate with new partners, even in a world where most of the interactions are one-shot. Across simulations, the proportion of interactions that are one-shot varied from 10% to 90%. (Even the lowest base rate of 10% probably overestimates the percent of one-shot partners experienced by hunter-gatherers, who lived in small interdependent bands and whose extra-band encounters were primarily with people from neighboring bands.) Each new partner comes with a number—a cue summary—that serves as a hint to whether an agent's interaction with that partner will be one-shot or repeated. The cue summaries are never perfect predictors: they are drawn from one of two normal distributions (one-shot vs. repeated) that overlap by either 13%, 32%, or 62%. Differences in how discriminable the cue summaries are accounted for < 1.2% of the variance in the magnitude of evolved cognitive variables, so results mentioned below are averaged over these three parameter values.

In one set of simulations, agents evolve a decision threshold determining how strong cues that the interaction is one-shot must be before the agent defects. Selection favored a threshold of evidence so high that most interactions were classified as repeated, triggering cooperation. In other simulations, the agents are Bayesians who develop rational beliefs by integrating (i) cues that the given interaction is one-shot, with (ii) the base rate of one-shot interactions in the population. What evolves is a regulatory variable that determines the probability the agent will cooperate *given its rational belief that the interaction is one-shot*. Consider a world in which the base rate of one-shot interactions in the population is very high—50%—with modest gains in trade ($b{:}c = 3{:}1$) and repeated interactions of 10 rounds on average. Selection favored designs with a ~90% probability of cooperating when the agent (rationally) believes the interaction is most likely one-shot (~70% probability when repeated interactions average five rounds). This probability was still high (~80%) when the base rate of one-shot interactions in the population was even higher—70%—and it evolved to be near 100% for gains in trade ≥ 4:1. With higher gains in trade and/or repeated interactions with more rounds, agents evolved a strong motivation to cooperate given a one-shot belief even when 90% of interactions in the population were one-shot.

The simulations with Bayesian agents are particularly apt because most subjects who cooperate in experimental economics games say they believed the experimenter's claim that they would be participating in one-shot interactions. The results show that natural selection can favor a disposition to start out cooperating, even in people who believe an interaction is most likely to be one-shot. No group selection is needed.

21. Partner Control, Reputation, and Punishment

Does a person's reputation affect how likely you are to cooperate with them? And if so, how long is the shadow of reputation? Strong reciprocity models based on group selection emphasize the use of punishment to maintain group norms of cooperation. The shadow of reputation is long in these models: individuals exclude norm violators from the benefits of reciprocal cooperation, whether the violator defected on third parties or oneself. In contrast, partner control models based on individual selection suggest that decisions to trust and cooperate will be most influenced by how the other individual has treated *you*. When you have no past interactions with a partner, information about that partner's willingness to defect on third parties may be used as a cue for deciding whether to trust a partner to cooperate with you.

To test these alternatives, Krasnow et al. (2012) gave people information about their partners' reputations before a two-round trust game. In one version of the experiment the reputational information was the partner's behavior in prior prisoner's dilemma games with the subject and with third parties; in another it was the partner's responses to moral dilemmas involving cheating. Results were the same: the shadow of reputation was short, and information about third-party norm violations regulated decisions to trust a partner only when subjects had not interacted with that individual prior to making a decision. When subjects learned that their partner had cooperated or defected on *them*, that was the sole factor regulating their decisions to trust and their subsequent decisions to cooperate or defect. In contrast to the predictions of group norm maintenance models, whether the partner had defected on third parties had no effect whatsoever once subjects knew how their partner had treated them.

Having defected on third parties had no effect on decisions to punish, either. There is increasing evidence that anger is the expression of a neurocomputational system that evolved to bargain for better treatment, by threatening to either inflict harm (punishment) or withdraw cooperation (Sell et al., 2009, 2017; Tooby et al., 2008). Deploying these costly tactics makes sense only when you plan to keep the relationship but reform the partner: partner control. Consistent with this view, Krasnow et al. (2012) found that when people punished their partner's defection, they were vastly more likely to cooperate with them on the next round than when they had left the defection unpunished—a pattern that makes sense if punishment is deployed as a way of bargaining for better treatment when you plan to continue the relationship. (Punishing the partner for defecting would be a waste of effort if you planned to abandon the relationship anyway.) By contrast, strong reciprocity/group norm maintenance theories predict that (i) punishment will be directed toward norm violators whether they defected on the punisher or someone else and (ii) the norm violators will be *excluded* from subsequent cooperative interactions. The results reported by Krasnow et al. (2012) suggest that motivations for punishing defectors evolved to maintain dyadic reciprocation, not to exclude norm violators from a cooperative group.

22. Partner Choice, Virtue Ethics, and the Evolution of Morality

From Aristotle and Confucius to the present, moral character has been a central concern of virtue ethics (e.g., Aristotle, 2005/350); McCloskey, 2006; Runes, 1983; Swanton, 2003).

Views about which traits are virtuous differ, but when there is a record of what people in a given culture find virtuous, some traits recur across cultures and time: generosity/

benevolence/kindness/compassion; honesty/integrity/loyalty; righteousness/justice/reciprocity; knowledge/skills/excellence; temperance/prudence/frugality; courage/bravery/strength. Taking virtuous actions is not sufficient to make you a virtuous person—an individual may pursue a generous, just, or honest course of action for self-serving, cowardly, or prudential reasons. A virtuous person is one who has cultivated generosity, honesty, justice, and other virtues as part of their character. In this view, an honest person behaves honestly because she values honesty: the fact that honesty demands it would be, for her, a strong reason to tell the truth (Hursthouse & Pettigrove, 2016). But why should this distinction between action and character ever matter to us as products of natural selection? Natural selection favors cognitive designs because of their reproductive consequences. The utilitarian consequences of cooperation—the fact that partners provided food, favors, and help—drove selection for adaptations that cause and regulate cooperative behavior. So why should we care whether a person helped us for self-serving reasons or as an expression of their character? Partner choice theories suggest an answer.

Adaptations for partner choice are expected when there is a "biological market"—that is, when there are many potential partners available and, therefore, competition for the best ones (for reviews, Barclay, 2015, 2016; Baumard et al., 2013). Social arrangements that allow one to break links with defectors elicit higher levels of cooperation from people (Fehl et al., 2011; Rand et al., 2011), but partner choice is as much about choosing good partners as avoiding cheaters. When individuals can choose among partners who vary in cooperativeness, models show that a positive feedback loop selects for high levels of cooperation and choosiness (e.g., McNamara et al., 2008).

23. The Origins of Fairness

To simplify their models, economists sometimes assume that people reason in accordance with game theory, with the goal of maximizing their monetary payoffs. If we take this as a model of human motivation, then the results of dictator and ultimatum games are puzzling. In the one-shot dictator game (DG), the proposer is given money (e.g., $10) and told she can keep all of it or give any fraction to another (anonymous) person. But instead of keeping the full $10, most people divide it, and many divide it equally. Similar results are found with the ultimatum game (UG): a dictator game in which the other person can respond. The responder can either accept or reject the offer made by the proposer. If the responder accepts the offer, the money is divided accordingly; if the responder rejects the offer, proposer and responder get nothing. That is, there are gains in trade only if both parties agree to the split offered. If both parties were reasoning in accordance with game theory, however, the proposer would offer the smallest amount possible (to maximize her payoffs) and the responder would accept (because something is better than nothing). Precisely that pattern evolves in evolutionary models when there is no partner choice or when the cost of switching partners is high (André & Baumard, 2011; Debove et al., 2015). Yet many proposers offer 40–50%, and these offers are accepted (Güth & Kocher, 2014). Moreover, many responders reject low offers (~20%), thereby paying to punish the proposer. In the literature, giving half is considered "impartial" (it does not favor self over other) and "fair" (benefits are distributed in proportion to the costs incurred to produce them—in most UGs the money distributed is not earned by either party; it is a windfall from the experimenter).

Fair, impartial divisions in these games are very common in advanced market economies. How often such divisions are made in small-scale societies is predicted by the degree to which their local economy is integrated with mass-market economies (Henrich, Ensminger, McElreath, et al., 2010). Market economies provide evolutionarily unprecedented opportunities for partner choice: the division of labor means that people with diverse skills are offering a wide variety of goods and services to one another. This raises the question: Can partner choice explain the evolution of fairness? Baumard et al. (2013) argue that moral intuitions about fairness and impartiality evolved via partner choice in biological markets.

If they are correct, then the proportion of the pie offered and accepted in UGs with many rounds will shift with changes in the supply and demand for partners. It does. Theoretical models (André & Baumard, 2011; Debove et al., 2015) and behavioral experiments (Debove et al., 2015) yield similar results. In experiments by Debove and colleagues, competitive altruism emerged when there were more proposers than responders—that is, when responders had more "outside options" and, therefore, more bargaining power. As proposers competed to be chosen by the responder, the average offer accepted climbed to ~90% of the pie. Conversely, offers fell to ~25% of the pie when responders outnumbered proposers; proposers offered less and less when responders were competing for an offer. But divisions stabilized ~50% when participants could change roles after each round, an arrangement that equalized the outside options of proposers and responders. When proposers were keeping > 50%, responders had an incentive to become proposers; when they were keeping < 50%, proposers had an incentive to become responders. This led to an equilibrium in which the pie was equally divided. "Market forces" were sufficient to create these outcomes; participants were never told the mix of proposers and receivers (Debove et al., 2015). In other models, they found that dominant individuals could not leverage their ability to coerce when weaker individuals frequently encountered alternative partners—partner choice created an equilibrium near 50% (Debove et al., 2015).

Ability to Provide Benefits

In the experiments just described, the resources distributed were a windfall from the experimenter. But in real life, effort, skill, and social influence play a role in the production of resources. People vary in these "factors of production"; all else equal, people with higher ability to provide benefits make better cooperative partners. Cues that predict a partner's ability to provide benefits do matter in partner choice (for review, see Barclay, 2015, 2016)—not just modern ones, such as wealth and income, but ancestrally reliable cues, such as health, attractiveness, social status, strength (in men) and, interestingly, judgments of how productive a group member the partner would be if he or she "lived 100,000 years ago, when humans had to hunt or gather food and find or build shelter" (Eisenbruch et al., 2016). Using responders whose faces had been pre-rated on these dimensions, Eisenbruch and colleagues conducted a series of one-shot UGs where proposers could see a photo of each (same-sex) responder's face. For each face, participants said how much they would offer as proposers and how much they would demand as responders (strategy method). Both sexes were more generous toward responders who looked healthier, more attractive, higher status, stronger (men only), *and* more productive as hunter-gatherers. Men in particular behaved as if their offer was an opening bid to attract high quality long-term

cooperative partners: They made higher offers and demanded less in return from men rated higher on traits that predicted ancestral productivity—effects that remained even after controlling for how much reciprocity they expected in return. This strategy was costly: the more sensitive men were to these traits, the less they earned. Raihani and Barclay (2016) show that when fairness is held constant, people prefer partners whose ability to (stably) provide resources is higher.

It may seem odd that knowledge, skills, temperance, and prudence—factors that increase an individual's economic productivity—are often viewed as *moral* virtues. But if moral virtues reflect properties that make one a good cooperative partner in the eyes of others, this makes sense: factors that increase your own productivity also increase your ability to provide benefits to others. This may also explain why excellence at rare skills—"virtuosity"—is often considered a virtue. Honing unusual skills that others value—especially ones that shower positive externalities on others (Tooby & Cosmides, 1996)—increases one's value as a cooperative partner.

Willingness to Provide Benefits

An individual's ability to provide benefits is little use if they are unwilling to share them. Not surprisingly, generosity is a highly valued trait in potential cooperative partners (Barclay, 2015), and more generous partners are preferentially chosen (Barclay & Willer, 2007). In Eisenbruch et al.'s (2016) study, men and women offered more to partners who looked more kind, cooperative, and trustworthy. For men, these faces elicited a boost in generosity beyond what could be explained by their expectations about how much these partners would demand as responders or offer them as proposers. As with productivity, men incurred a one-shot cost by being sensitive to these traits: They behaved as if they were trying to attract long-term cooperative partners who could be counted on to help them.

The fact that a reputation for generosity is valuable provides a second reason to expect generosity even in one-shot, anonymous interactions. Adults and children (ages 5–8) are more generous when their actions might be seen than when they are private (Barclay, 2015; Kraft-Todd et al., 2015; Shaw et al., 2014; Leimgruber et al., 2012). When people know that third parties will see what they gave in a prior, one-shot game, partner choice makes them "competitively altruistic": they give more when the third party can choose their own partner than when partners will be assigned randomly (Barclay & Willer, 2007; Sylwester & Roberts, 2013).

24. Moral Action versus Moral Character

Partner choice creates incentives to signal that you are a good cooperative partner, whether this is true or not. This favors adaptations in choosers for distinguishing opportunistic generosity from "true" generosity—generosity that can be relied on even when short-run incentives favor defection. The same is true for fairness: A person might divide resources with you impartially when you have many outside options but take advantage of their superior bargaining power when circumstances allow. How can a chooser know who to trust?

In the coevolutionary arms race between deceptive signalers, truthful signalers, and receivers who benefit from accurate information, the common interests of honest signalers

and receivers tends to select for signals that are honest because they are difficult to fake (e.g., cues of health as indices of ability to produce resources) or costly to the signaler, either energetically or socially (Higham, 2014). Forgoing other cooperative relationships to invest in a particular partner is a social cost, for example; forgoing resources that the signaler could consume or use to extract immediate gains from reciprocity is an energetic cost. In laboratory experiments, costly signals inspire trust. Individuals who take the opportunity to donate to a charitable organization were trusted more by exchange partners and, in fact, proved more trustworthy (Fehrler & Przepiorka, 2013). People who sacrifice more to benefit the chooser are preferred to those who sacrifice less to deliver the same benefit—Lim (2010) found that partners who sacrificed more were preferred even when the partner who sacrificed less had given three times as much (see also Raihani & Barclay, 2016). Among Martu hunters in Australia, centrality in the cooperative hunting network (a measure of partner preference) is better predicted by the proportion of a catch a hunter shares than by his hunting skill (Bliege et al., 2015). Bliege et al. (2015) argue that "pecuniary distancing" by successful hunters—taking less of the catch yourself, giving others control over how it is distributed—serves as an honest signal of cooperative intent.

A reputation for impartial equity is so important to children that they will destroy resources that they could otherwise have consumed rather than appear to favor themselves over others (Shaw & Olson, 2012). Everett et al. (2016) argue that people with an aversion to using a human being as a mere means to an end may be preferred as cooperative partners because they will more easily resist temptations to inflict harm to benefit themselves. Using trolley problems—moral dilemmas in which sacrificing one person's life will save five—they compared reactions to people who made deontic versus consequentialist judgments. On the footbridge version, where the man sacrificed is being used as a means to achieve an otherwise desirable consequence, those who made deontic judgments (i.e., thought it wrong to push the man onto the tracks to stop the trolley) were perceived as more moral and trustworthy; they were also trusted more in economic games and preferentially chosen as cooperative partners. Tellingly, deontic judgments did not elicit these positive character attributions when the man's death was a side effect of switching the train onto another track. The key variable was refusing to use the man as a means to an end.

In other words, the evolution of cooperation by partner choice provides a scientific rationale for distinguishing between actions that benefit others and moral character (Baumard et al., 2013). Generosity, honesty, integrity, loyalty, reciprocity, and justice are cited as moral virtues across cultures and time. Those who exhibit these virtues even when they have short-term incentives to act otherwise make better long-term cooperative partners, and, because they are more likely to be chosen as partners, their own well-being may be enhanced in the long run. In this way, good moral character could contribute to *eudaimonia*, a life in which the individual flourishes (Aristotle, 2005/350 BCE).

25. Different Sharing Rules Are Triggered by Luck versus Effort

What is "fair"? Dividing a resource equally or equitably? In reviewing laboratory and cross-cultural evidence, Baumard et al. (2013) find a widespread preference for equitable distributions, in which benefits are distributed in proportion to effort and talent. In distributing the benefits of cooperation among third parties, children age 3–5 take merit into account—that

is, how much work each party contributed to producing the benefit (Baumard et al., 2012; Liénard et al., 2013). Sensitivity to merit is not restricted to children from WEIRD cultures (Western, Educated, Industrialized, Rich, Democratic; see Henrich et al., 2010). It is found in 5-year-old Turkana children raised as pastoralists in northern Kenya (Liénard et al., 2013). How much work a child did relative to a cooperative partner matters even when preschoolers are distributing benefits between self and other (Hamann et al., 2014; Kanngiesser & Warneken, 2012). Why, then, is it common for adults from advanced market economies to divide resources equally in dictator and ultimatum games? Do they have a general preference for equality over equity?

No. In most DG and UG games, the resource proposers are dividing is a windfall from the experimenter. When proposers have to work to earn the resource they are then invited to divide, they share very little (List, 2007; Oxoby & Spraggon, 2008). And when DG proposers are given two possible actions—to give *or* take from the recipient—windfalls elicit behavior that is not impartial, equal, or "prosocial" (List, 2007). When endowments were earned, most proposers (66%) gave nothing to the recipient *and* they took nothing from the recipient. But when endowments were a windfall, 30% gave nothing to the recipient and > 55% of proposers *took* money from the recipient (List, 2007). Why is the same amount of money, in each case provided by the experimenter, distributed differently when it is a windfall rather than earned? Is the key variable effort expended or something else?

In a classic study of Ache foragers in Paraguay, Kaplan and Hill (1985) found that the same individuals in the same culture applied different sharing rules to meat and honey than to gathered plant foods. Meat and honey were shared widely in the band—they were communally shared (Fiske, 1991), according to a rule approximating Marx's claim that hunter-gatherers share "from each according to their ability, to each according to their need." This was not true, however, for most of the gathered foods. These were shared within the family or with specific reciprocity partners. Effort was required to acquire all of these resources; foraging risk was the variable that explained which sharing rules were used for each resource.

Hunting is a high risk-high payoff activity. Behavioral ecologists studying tribal societies find that hunters come back empty handed on more than half of their hunting trips (Kaplan et al., 2012). These reversals of fortune apply across skill levels: effort is not sufficient to ensure hunting success. When hunters do succeed in killing an animal, there is often more meat than one family can consume. Keeping this extra meat for future consumption is not practical, because of decay and the energetic costs of transport for semi-nomadic people. So hunter-gatherers store this extra food in the form of social obligations (Cashdan, 1982). They buffer high variance in foraging success by pooling their risk (Cashdan, 1982; Kaplan & Hill, 1985). My family eats today, even though my hunt failed and yours succeeded, because you share your catch with me; tomorrow, when you fail and I succeed, your family still eats because I share with you. Honey is shared widely for the same reason: The payoff is large, but there is high variance due to luck in finding and acquiring it. Gathered plant foods are different. Their caloric density is usually lower than for meat and honey, there is little variance in gathering success, and what variance exists is largely due to effort expended, not luck. Under these circumstances, risk-pooling offers no advantages. These low risk-low payoff foods are the ones shared within the family or with specific reciprocity partners (Kaplan & Hill, 1985).

Evoked culture or cultural transmission? This pattern—band-level sharing for high risk-high payoff foods, reciprocal sharing for low risk-low payoff foods—is typical for hunter-gatherers. But why? Is it because they have inherited packages of cultural norms that gradually accumulated over time because they worked well in this ecology? Or does this cultural pattern exist because our minds have (at least) two different evolved programs, each equipped with different sharing rules? In this view, cues of high variance activate different sharing rules than cues of low variance. When variance is high, this triggers an evolved program that generates the intuition that the lucky should share with the unlucky; when variance is low, the evolved programs activated generate the intuition that you have no obligation to share outside the family, except, perhaps, with specific social exchange partners. This second possibility is what Tooby and Cosmides (1992) call "evoked" culture: the cultural pattern is evoked by the situation—that is, it emerges because the mind is designed to activate different programs in response to cues of different ancestral situations. An evoked culture explanation predicts that cues of high versus low variance will activate different sharing rules in humans everywhere, not just among hunter-gatherers. Explanations that invoke the accumulation of norms by success-biased cultural transmission do not predict this cue-activated pattern (e.g., Henrich, 2015; Richerson & Boyd, 2006).

When history and ecology differ across cultures, success-biased cultural transmission should create different packages of norms, each appropriate to the local culture. In advanced market economies, we forage at grocery stores where variance due to luck is low, we live in family units rather than bands, and when we buy food, most of it is shared within the family. We are WEIRD people, who should have different sharing norms than hunter-gatherers. Cultural evolutionists have argued that WEIRD people are unusual compared to the rest of the species (Henrich et al., 2010) and that our sharing norms are very different from those found in small-scale societies (Henrich et al., 2005).

Yet WEIRD people respond to cues of high versus low variance just like hunter-gatherers do. For example, holding effort expended constant, Japanese and American college students were more willing to share money acquired via a high variance process than a low variance one; moreover, the effect of high variance was independent of individual differences in the students' ideologies about the just distribution of resources (Kameda et al., 2002).

An ingenious test of the evoked culture prediction was conducted by Kaplan et al. (2012). By creating a foraging game in a virtual world, in which each (anonymous) subject has an avatar with a "food pot," Kaplan and colleagues showed that WEIRD students from southern California immediately detect which of two foraging patches has high versus low variance, and respond like hunter-gatherers. When they successfully gather food, they can choose to deposit the calories in their own pot or in pots of other avatars (calories in your pot determine earnings). When subjects foraged and caught food on the low variance patch, they did not share it—they usually put all the calories in their own pot. But when they foraged on the high variance patch, lucky subjects shared with unlucky ones by putting calories from their catch into the pots of other avatars. Experiencing the high variance patch elicited more sharing from the first round of the game (with <2 minutes of foraging experience). Spontaneous reciprocal sharing emerged and increased over 20 rounds of foraging, but only when subjects foraged on the high variance patch. By the end of the hour,

the difference in calories shared was dramatic: 50 times higher for food caught on the high variance patch compared to the low variance one.

If alternative evolved systems regulating sharing are triggered by perceptions of luck versus effort, we should see the fingerprints of these systems in moral intuitions about public policy. Social welfare programs are designed to help people in need, but need due to bad luck should trigger different moral intuitions than need due to low effort (Petersen, 2012). Welfare programs have more citizen support in Denmark than the US, and stereotypes of welfare recipients differ accordingly: Danes are more likely to attribute their need to bad luck, Americans to laziness. Attributions of laziness should trigger moral intuitions about withdrawing cooperation from cheaters, not sacrificing to provide them with benefits. Through nationally representative surveys, Aarøe and Petersen (2014) tested the evoked culture view experimentally by varying perceptions about a man who is currently on social welfare. When participants were asked to imagine a man who had a regular job in the past, suffered a work-related injury, and is motivated to work again, opposition to social welfare benefits decreased in both countries—and to the same level. When he was described as a healthy man who has never had a regular job and doesn't want one, opposition increased in both countries—again to the same level. Two sentences portraying need due to bad luck versus low effort changed intuitions about which sharing rules were most appropriate and eliminated the difference between Danes and Americans in attitudes about welfare benefits.

This research is directly relevant to moral pluralism. Perceptions of luck versus effort trigger very different intuitions about what kind of sharing is morally appropriate. This variation is highly systematic, however, because it is caused by the activation of different evolved systems in the mind.

26. Cooperation in Groups: Collective Action, Free Riding, and the Evolution of Punishment

Collective action is when three or more individuals cooperate to achieve a common goal and then share the resulting benefits. Among hunter-gatherers, cooperation via collective action is common in intergroup aggression, hunting (especially big game), shelter building, and, more generally, in any context where a resource is more easily harvested or produced by the coordinated cooperation of several individuals.

When collective action is seen in other animals—bees, ants, wolves—those cooperating are usually kin; selection can favor adaptations for helping kin, even when kin do not reciprocate (see §5–11, especially §7). The fact that we humans easily engage in collective action with individuals to whom we are not genetically related is a zoologically unusual feature of our species. The list of species who do the same may eventually grow, but aside from humans, the only uncontroversial examples are chimpanzees and dolphins—species in which males coordinate to attack members of other groups and defend against similar attacks (e.g., Wrangham & Peterson, 1997). A fight is a conflict between two individuals, but warfare is a conflict between two groups of individuals, each of which must coalesce and function as a cooperative unit.

Coalitional aggression is found among hunter-gatherers (Keeley, 1996)—more often as raids than battles—and the fact that chimpanzees also engage in intergroup conflict suggests that the selection pressures it creates have been shaping our psychology for perhaps

6 million years (i.e., since hominins and chimpanzees had a common ancestor). A large literature on the psychology of "us" versus "them" shows that it is remarkably easy to elicit identification with a group and ethnocentrism, including a pattern of ingroup favoritism and outgroup derogation that includes moral attributions (for reviews see Mackie et al., 2008; Sidanius & Pratto, 1999). Killing ingroup members is frowned upon (although sometimes tolerated) in all human societies (e.g., Boehm, 2001, 2012; Wrangham et al., 2006; Wrangham, 2019), but the ethnographic and historical record is replete with examples of people dehumanizing, stealing from, and killing outgroup members—an activity that is sometimes celebrated (Pinker, 2011; Tooby & Cosmides, 2010). Warriorship contributes to men's status in small-scale societies (Chagnon, 1988; von Rueden, 2014), and, in many contexts, men respond to outgroups and intergroup conflict differently than women do (McDonald et al., 2012; Sidanius & Pratto, 1999). In one laboratory example, intergroup competition led men to identify more with their ingroup and make more generous contributions to their group in public goods games; the same manipulation had no effect on contributions by women (Van Vugt et al., 2007). That virtues in war—bravery, courage, loyalty, and strength—were often mentioned as moral virtues in past societies is consistent with this feature of our species' evolutionary history.

Data from across the behavioral sciences suggest that the human mind includes a *coalitional psychology*: a set of neurocomputational programs that evolved to regulate within-group cooperation and between-group conflict (Kurzban et al., 2001; Pietraszewski et al., 2014; Sidanius & Pratto, 1999; Tooby & Cosmides, 2010; Tooby et al., 2006). These programs regulate motivations for participating cooperatively in coalitions, policing their boundaries, and interacting with outgroups. Models of selection pressures for initiating coalitional aggression and defending against it suggest that intergroup conflict activates moral intuitions, judgments, and sentiments very different from those regulating collective action for other purposes (see, e.g., Choi & Bowles, 2007; Sidanius & Pratto, 1999; Tooby & Cosmides, 1988). To date, these differences have not been an active area of research (but see McDonald et al., 2012; Van Vugt et al., 2007).

Earlier we discussed why cognitive adaptations for social exchange cannot evolve unless they include mechanisms for directing benefits to other cooperators and away from cheaters. A similar—but more difficult—problem is common to collective action in all domains. Collective action often produces public goods: benefits that will accrue to everyone in the group, whether they contributed to producing the good or not. (Group defense is a common example.) When this is true, there are incentives to *free ride*: to contribute less than other group members to producing the common goal (or contribute nothing at all; Olson, 1965). Strategies that motivate free riding will have higher fitness than strategies that motivate contributing to the collective action, thereby selecting against cognitive systems that motivate contributing to collective actions. Adaptations for collective action cannot evolve unless the fitness payoffs for contributing exceed those for free riding. Pure public goods are an extreme point on a continuum, where ostracism from the group is the only way to exclude a free rider. The practicality or cost of decreasing the access free riders have to the fruits of collection actions will depend on the good produced, its location, and how it is shared.

Partner choice in a biological market is a solution that can work well for two-person exchange (see §22–24). As long as better partners are available, you can withdraw from your

current partner and choose another. Group cooperation is a different matter: You cannot withdraw your cooperation from the free rider without leaving the group. That entails leaving the high contributors as well, thus forgoing the large benefits that can be attained by collective action but not individually (such as hunting large game or defending against raiding groups). Partner control is a better option: negative sanctions, including punishment, can incentivize the free rider to contribute more in the future.

Results from economic games confirm the importance of punishment in sustaining group cooperation. In a public goods game (PGG), each player is given tokens (worth money) that they can keep or invest in a common pot. Any tokens contributed to the pot are multiplied by the experimenter and then distributed equally to each player. The multiplier is chosen so that each player makes more if they keep their tokens but everyone else contributes. But since everyone has the same incentive to free ride, game theory predicts no one will contribute.

Usually people start out contributing 40–60% of their endowment to the common pot. But when there is no mechanism for punishing free riders, the amount players contribute to the common pot decreases over rounds, until the contributions dwindle to almost nothing. This is not because people failed to understand the game at first; when they are paired with a new set of players, they once again contribute 40–60% of their endowment and, over rounds, the same thing happens: contributions dwindle.

Behavior is dramatically different, however, when the game allows people to punish other players. Even though punishment is personally costly (you might pay one token to deduct three from another player), some people punish. Instead of dwindling over rounds, contributions to the common pot remain high and sometimes increase (Fehr & Gachter, 2000; Masclet et al., 2003; Yamagishi, 1986). Masclet et al. (2003) show that under-contributors who are sanctioned increase their contributions on subsequent rounds, and sanctions need not be monetary to work; sending "disapproval points" increases contributions from free riders, as long as most of the other players are sending them. Two factors independently predicted how much high contributors punished under-contributors: how much less the player contributed compared to (i) the average amount contributed in that round and (ii) how much the punisher had contributed in that round. Other methods also show that those who contribute more feel more punitive sentiments toward free riders (Delton & Cimino, 2010; Price et al., 2002). Interestingly, people appear to engage in preventive moralization—they take a dim view of people who do not wish to participate in a collective action, even ~~when they~~ are willing to forgo the benefits it produces. The decision to opt out elicits negative moral evaluations, with worse views when the collective action produces public goods than ones that can more easily be restricted to participants. Experiments show that these negative moral evaluations are driven by the subject's perception of how likely the target is to free ride (Delton et al., 2013).

According to some models, group selection is necessary to account for the evolution of group cooperation and punishment, because paying to punish free riders will be selected against (Bowles & Gintis, 2013). In this view, those who contribute without punishing are second-order free riders: everyone reaps the benefit of greater future contributions from free riders, but only the punisher incurred the cost of punishing them. But agent-based simulations making more realistic assumptions show that adaptations for group cooperation

and punishment easily evolve via individual selection when (i) punishment is probabilistic (each agent has a variable, which can evolve, representing the probability the agent will punish any given act of free riding), (ii) punishment raises the probability that the free rider contributes in the future—but only when in the presence of the agent who did the punishing; and (iii) agents participate in multiple, partly overlapping groups (Krasnow et al., 2015). Under these conditions, punishers disproportionally reap the benefits of increased cooperation from the free riders they punished, so there is no second-order free rider problem. Cognitive designs with a positive probability of punishing emerged, and their presence selected for organisms designed to cooperate by default (i.e., without being induced to do so by punishment). In this social ecology, high levels of group cooperation emerged, in groups both small and large—up to 25 members (Krasnow et al., 2015).

27. What Criteria Does the Mind Use to Categorize Someone as a Free Rider?

We have been talking about free riders as if this were an unproblematic concept, but it is not. It might seem that contributing less than others to the collective action would be sufficient for classifying someone as a free rider, but an evolutionary perspective suggests otherwise.

A categorization system that used level of contribution as its sole criterion for recognizing free riders would generate no misses but many false alarms (cooperators incorrectly categorized as free riders). This is because every individual endowed with neurocognitive mechanisms that reliably motivate conditional cooperation will sometimes fail to contribute to a collective action due to bad luck, injury, errors, or accidents. Categorizing these conditional cooperators as free riders, to be punished or excluded, will trigger cycles of mutual defection, preventing cooperators from harvesting the benefits of repeated mutual cooperation that occur when conditional cooperators correctly recognize one another (see §20, on the evolution of generosity). Without these benefits, decision rules that motivate cooperation are outcompeted by those responsible for free riding, and collective action disappears from the population. This makes false alarms very costly fitness errors for organisms designed to cooperate conditionally. These considerations suggest that selection will favor a free rider categorization system biased toward minimizing false alarms.

The Anatomy of a Moral Concept

Research on social categorization shows that memory errors can unobtrusively reveal how the mind is classifying people (individuals who the mind has sorted into the same category are more easily confused than those sorted into a different category). Using this measure, Delton and colleagues showed that under-contributing to a collective action does not, by itself, result in an individual being classified as a free rider (Delton et al., 2012).

In each of six experiments, subjects watched a scenario about eight survivors of an airplane crash on a desert island. These men (the targets) agreed to forage for food and bring whatever they found back to camp to share with each other and those who had been injured in the crash. Subjects then saw what each man did on five different days of foraging (his face paired with a description of the event). After this, there was a surprise recall task:

subjects were asked who did what for each of the 40 events. In the first two studies, every target failed to bring back food on two days, but for different reasons (e.g., four found food but ate it themselves, the other four found food but accidentally lost it). In other experiments, some men found (and contributed) food on more days than others did. By seeing which targets subjects confuse when they make mistakes, one can infer how their minds classified them[16] (if at all). Afterwards, subjects rated each man on morally relevant character traits and his desirability as a cooperative partner.

The results were clear. People who tried to contribute but failed were not categorized as free riders—even when they contributed less than others. To be categorized as a free rider, the target had to under-contribute in ways suggesting an exploitive intent—that is, a motivation to benefit from the collective action without incurring the costs of contributing to it (e.g., eating the food he found instead of sharing it with the group; taking a nap instead of trying to find food). Memory errors showed that targets with exploitive intent were sorted into a distinct mental category from those who tried but failed to contribute due to accident or bad luck, and the individuals who intentionally under-contributed were evaluated more negatively on morally relevant dimensions (they were seen, e.g., as less trustworthy and more deserving of punishment).

The results could not be accounted for by a domain-general process that sifts for any behavioral difference between targets and uses it to categorize them. In a control condition, targets who tried to contribute but failed in one of two distinct ways (four failed to find food, four found food but accidentally lost it) were not sorted into two distinct mental categories.[17] Yet these were the same actions that had elicited categorization in the prior studies, when the other targets were free riders. Another study varied the targets' productivity. Those who tried to contribute but produced less than others were not categorized as free riders (just as less competent).

Although these results suggest a domain-specialized system for categorizing free riders, they could also be accounted for if the mind has criteria for distinguishing those who violate moral rules from those who do not. To test this "moral violator" counterhypothesis, Delton et al. conducted parallel experiments in which some targets were free riders and others committed a different kind of moral violation. The results showed that subjects spontaneously distinguished free riders from other kinds of moral violators. The most interesting test involved a very subtle difference. As in the other memory confusion experiments, subjects saw that everyone who had agreed to participate in the collective action contributed resources on three days. But on the other two days, subjects saw that some targets consumed a resource they had promised to contribute to the group (free riders), and others stole a resource owned by the group. Every one of these targets was intentionally violating a moral rule—*and* illicitly taking a benefit for themselves that was obligated to the group. Nevertheless, subjects sharply distinguished them, as revealed by the categorization measure and the response and character ratings gathered subsequently.

That the mind slices the moral domain so thinly is remarkable: stealing a resource from the group and consuming a resource promised to the group are so similar that many approaches to moral psychology would not distinguish them. These experiments suggest that our minds really are prepared to notice and remember which individuals are free riders on collective actions, making very subtle distinctions between free riders and people who commit other kinds of moral violations.

28. Beyond Utilitarianism and Deontology

Natural selection has produced computational systems equipped with moral concepts, inferences, judgments, and sentiments. But we hope these examples have demonstrated that the moral psychology of our species cannot be easily captured by normative theories organized around single principles, whether the principle is utilitarian (e.g., "maximize aggregate welfare") or deontological (e.g., "share resources impartially"). What is known about how natural selection has shaped moral cognition is more consistent with normative frameworks that embrace a pluralistic conception of moral rules and principles.

It has been argued that variation in commonsense moral intuitions undercuts the normative projects of ethical intuitionists and moral sentimentalists. But research demonstrating moral diversity does not necessarily undermine these projects if the variation is systematic. As the research discussed in this chapter shows, moral intuitions vary across domains of social interaction, as well as situations, relationships, individuals, and cultures. But we have also attempted to show that at least some of this variation is systematic. Many phenomena that seem contradictory or irrational resolve into patterns as the evolved architecture of the mind is uncovered.

If the view emerging from evolutionary psychology is correct, then our intuitions about how we ought to treat others and how they ought to treat us are produced by a number of different evolved systems, each specialized for regulating a different class of social interactions. Understanding how these domain-specialized computational systems parse and react to the social world could illuminate many issues in moral epistemology. This knowledge could provide a principled basis for delineating different moral domains, each with distinct duties, principles, and concepts of what is good and right. It provides practical information about how to reframe situations to activate alternative evolved systems, in ways that promote behavior that realizes a normative moral ideal. And it is useful for any philosopher who finds it important to articulate ethical principles that are likely to be accepted by animals with minds like ours.

Notes

1. Every feature of every organism is produced by the joint interaction of genes and the environment. Saying that a mutation produces a change in a design in no way denies (i) that development occurs, (ii) that certain environmental features are necessary for the design to develop in a particular way—which may include features of sociocultural environments, especially those that were reliably present ancestrally (e.g., nursing mothers; a community of language speakers), (iii) that the design may develop differently when those environmental features are absent or different, or (iv) that information from the current social and cultural environment influences how an individual behaves. (Indeed, research discussed in this chapter on the developmental cues used by the kin detection system illustrate all four points.) For explanations, see Tooby & Cosmides, 1992.
2. A tool or system has been designed when its features exist and were organized as they are because of their *functional consequences*. Natural selection designs adaptations in exactly this sense; so do people making tools (a tool has certain features because those features perform a function intended by its maker). Indeed, intentional design by an intelligent agent is a downstream consequence of natural selection designing our cognitive architecture.
3. Evolutionary psychologists emphasize hunter-gatherer life rather than more recent historical developments (such as permanent settlements, agriculture, or the industrial revolution) because computational systems capable of solving adaptive problems typically have a complex functional

design. Consider the eye, which is the front end of a computational system designed to use electromagnetic radiation to identify objects and their locations. The eye has a complex design that is functional: it has many parts that must work together in just the right way to focus and transduce light, the adaptive problem that selected for its design. The evolution of a complex functional system requires a number of separate allele fixations. That takes many more generations than a simple quantitative change, such as down-regulating the production of melanin in the iris (i.e., selection for blue eyes at high latitudes). Given a twenty-year generation time for humans, programs with a complex functional design would take thousands of years to evolve (Tooby & Cosmides, 1990b). It is unlikely that we evolved new programs to solve problems of social interaction that are unique to agriculture, given that half the human population was still hunting and gathering as recently as 5,000 years ago (farming first appeared in a few places ~10,000 years ago).

4. Many genes, working in concert with the environment, are necessary to produce every adaptation. To say that a mutation caused a difference in the design of the adaptation implies that at least one gene has changed. That mutation is a gene "for" the new design; the alternative allele—the one it replaces—is a gene "for" the original design. This language, which is standard in evolutionary biology, does not imply that complex adaptations are coded for by single genes.
5. In a diploid species, a given chromosome in a parent (and therefore any mutation on that chromosome) has a ½ chance of being passed on to a given offspring. For every two offspring produced, one will inherit a mutation from the parent. That probability is the same for every offspring produced, whether it is by the self or a sibling, so we can ignore this when considering how many copies of a mutation are produced by self versus sibling. What matters for this analysis is the probability that a mutation in self is also found in the sibling.
6. The *always cooperate* design will get positive payoffs for help given when ½ $B_{sib} > C_{self}$. But every time a resource or unit of energy expended on help is almost as valuable to self as to the sibling, this design will help. To the extent that you and your sibling are living in the same environment and have similar needs, there will be many cases like this. In each of these cases, your helping will cause a net decrease in copies of the mutation in the gene pool.
7. A common misunderstanding of Hamilton's rule is that individuals are designed to help full siblings because they "share half their genes" (known as the "fraction of genome fallacy"; Dawkins, 1979). In Hamilton's rule, $r_{self, kin}$ does not refer to the *fraction* of the entire genome shared by self and kin. It is the probability that self and a given kin member share *a given mutation* (one producing a Hamiltonian design for helping), regardless of how many other genes they may share in common. Although it is true that full siblings in a diploid species share half of their (nuclear) genes *on average*, with some sharing more and others less, that fact is irrelevant to the spread of a Hamiltonian mutation. The fraction of genome fallacy has led to incorrect inferences: e.g., kin selection does not imply that individuals will be more inclined to help people who share a larger fraction of their genome by virtue of ethnicity or any other factor. That is, kin selection cannot explain ethnocentrism or any other population-based social preference.
8. Random mutations are always occurring at low levels. By "universal," biologists mean that the design develops in everyone, except for the minute number of cases in which a mutation disrupts its development. Population genetic models suggest that disorders occurring at a rate of 1 in 1,000 most likely result from random mutations rather than being side effects of adaptations that are currently under selection.
9. When every member of the population has the Hamiltonian mutation, why don't organisms start indiscriminately helping non-kin (who, at this point, have the same design)? Remember that the Hamiltonian mutation codes for a motivation to help *kin* when $r_{self,kin}$ $(B_{kin}) > C_{self}$; it does not code for indiscriminate helping. A new mutation could arise that suppresses or alters the Hamiltonian one, producing individuals who help others regardless of kinship whenever $B_{other} > C_{self}$—Design #4. But the same logic applies: Given opportunities to help where $B_{sib} > C_{self} > ½ B_{sib}$, Design #4 helps siblings, thereby reducing its own replication relative to the Hamiltonian design, for the same reason that Design #2 does. When Design #4, which helps when $B_{other} > C_{self}$, helps non-kin—who have the Hamiltonian design—it reduces its own replication relative to the Hamiltonian design.

10. Being told by others is not a solution: this just pushes the problem one step back. The teller would have to know who is a *genetic* relative, how close a relative they are, and that it *matters* whether an individual is related by genes, marriage, or affinity. Even worse, misrepresenting this information is sometimes in the interest of the teller (e.g., mothers should want their children—whether full or half-sibs—to be more altruistic toward one another than they should want to be); see Trivers (1974) on parent-offspring conflict and the section below it on family dynamics. Other problems arise because kin terms are often used metaphorically to convey a close relationship (e.g., using "my brother!" when greeting a close friend) or to foster such relationships (e.g., a mother encouraging her child to address a close family friend as "Aunt Ellie," or referring to a stepsibling as "your sister").
11. When a lion eats a zebra, the zebra has increased the reproductive success of the lion and decreased its own reproductive success. But this effect was not by design (and, therefore, is not considered altruism in biology). Zebras have adaptations designed for escape, not for running into a lion's mouth. Note that a design feature can be altruistic in the biological sense without involving intentions, knowledge, or even behavior. There are, for example, trees that respond to their leaves being eaten by releasing volatile chemicals that are sensed by neighboring trees; on sensing these chemicals, the neighboring trees produce more toxins that are distasteful to leaf-eating herbivores. Producing volatiles and releasing them is an altruistic design feature.
12. Kin also engage in social exchange—indeed, it is expected for resources or actions whose costs and benefits fall outside the window in which kin-selected adaptations would motivate unconditional (i.e., non-compensatory) helping.
13. The rate of conversion from payoffs (calories, status, favors) to offspring is determined by the modeler in evolutionary game theory and by nature in natural selection. The psychology of individual organisms does not convert payoffs in calories, status, favors, and so on into cognitively represented estimates of their effects on future reproduction. (It may sometimes look that way, however, because evolved mechanisms for estimating the value of calories, status, favors, and so on can be expected to respond to ancestrally reliable cues of health, ovulatory status, kinship, caloric value, and other factors that affected reproduction ancestrally (see §8, "Estimating and Representing Benefits and Costs"). But notice the importance of cues (compared to effects on future reproduction): people enjoy artificial sweeteners knowing they have no caloric benefit; they enjoy sex while using contraception; they preferentially help stepsiblings with whom they were raised and are disgusted at the prospect of sex with them; and so on.) Very little is known about the psychology by which individual organisms estimate, represent, and compare payoffs within domains (e.g., alternative foods, or the value of a unit of food to self vs. other) and across domains (e.g., allocating time to eating versus romantic opportunities; see Roney & Simmons, 2017). Are calories represented by a specialized currency, different from that used for romantic opportunities, with pairwise "exchange rates" between domain-specialized currencies? Are payoffs in different domains translated into an internal lingua franca, a domain-general currency representing "satisfaction" or "utility"? These are interesting and important empirical questions that are unanswered at this time.
14. This may not be true for other species that engage in reciprocal behavior. Vampire bats, who transfer meals of foraged blood to one another, could have mechanisms specialized for representing volume of blood transferred; baboons could have mechanisms specialized for computing time spent grooming one another. Humans, by contrast, are capable of exchanging an open-ended set of tools, favors, and resources. For this reason, it was a prior prediction of the task analysis that algorithms for reasoning about social exchange in humans would extract an abstract representation of benefits and costs from concrete situations describing social exchange, and that procedures for detecting cheaters would operate on those representations (e.g., Cosmides, 1985; Cosmides & Tooby, 1989).
15. By concepts such as *obligation* and *entitlement*, we are not referring to the content of an obligation—the particular actions that people feel they are obligated or entitled to do vary hugely across cultures and history. *Obligation* and *entitlement* in the sense meant are concepts defined by their relationship to one another, to other inferences in the social exchange system, and to moral

emotions. For example, when agent 1 is *entitled* to receive *X* from agent 2, that implies that agent 2 is *obligated* to deliver *X* (research on social exchange shows that people spontaneously make this inference). It might also mean that if agent 1 takes *X* from agent 2, agent 2 will not punish agent 1 in response. The precise meaning of these evolved concepts is an empirical question; proposals about what they mean can be found in Cosmides (1985) and Cosmides and Tooby (1989, 2008a). Note that the concept of obligation used by cognitive adaptations for social exchange may not map onto colloquial concepts of obligation. Although the word "ought" is used in both circumstances, there are reasons (both theoretical and empirical) to expect the meaning of "ought" deployed by social contract algorithms to be different from the meaning of "ought" employed by a reasoning system specialized for interpreting and reasoning about precautionary rules (ones saying that a person "ought" to take a specific precaution when facing a particular hazard; Cosmides & Tooby, 2008b; Fiddick et al., 2000).

16. If two targets have been sorted into separate mental categories—male and female, for example—subjects will be more likely to make a within-category error than a between-category error (e.g., they will be more likely to misattribute something done by a woman to another woman than to a man). This pattern will emerge whether the subject is aware of classifying the targets or not.
17. For example, given that a man had failed to find food, subjects were just as likely to mistakenly attribute that event to a man who lost food as to one of the other men who failed to find food. That is, subjects were as likely to make a between-category error as a within-category error.

References

Aarøe, L. and Petersen, M. B. (2013). "Hunger Games: Fluctuations in Blood Glucose Levels Influence Support for Social Welfare," *Psychological Science*, 24 (12), 2550–2556.

———. (2014). "Crowding Out Culture: Scandinavians and Americans Agree on Social Welfare in the Face of Deservingness Cues," *Journal of Politics*, 76 (3), 684–697.

Adams, M. and Neel, J. (1967). "Children of Incest," *Pediatrics*, 40, 55–62.

André, J-B. and Baumard, N. (2011). "The Evolution of Fairness in a Biological Market," *Evolution*, 65, 1447–1456.

Aristotle. (2005/ 350 BCE). "Nicomachean Ethics," trans. W. D. Ross. Digireads.com.

Audi, R. (2005). *The Good in the Right: A Theory of Intuition and Intrinsic Value*. Princeton: Princeton University Press.

Axelrod, R. (1984). *The Evolution of Cooperation*. New York: Basic Books.

Axelrod, R. and Hamilton, W. (1981). "The Evolution of Cooperation," *Science*, 211, 1390–1396.

Barclay, P. (2013). "Strategies for Cooperation in Biological Markets, Especially for Humans," *Evolution & Human Behavior*, 34 (3), 164–175.

———. (2015). "Reputation," in *Handbook of Evolutionary Psychology* (2nd ed.), edited by D. Buss. Hoboken, NJ: John Wiley & Sons.

———. (2016). "Biological Markets and the Effects of Partner Choice on Cooperation and Friendship," *Current Opinion in Psychology*, 7, 33–38.

Barclay, P. and Willer, R. (2007). "Partner Choice Creates Competitive Altruism in Humans," *Proceedings of the Royal Society, London B*, 274, 749–753.

Baumard, N., André, J-B. and Sperber, D. (2013). "A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice," *Behavioral & Brain Sciences*, 36, 59–78.

Baumard, N. and Boyer, P. (2013). "Explaining Moral Religions," *Trends in Cognitive Science*, 17 (6), 272–280.

Baumard, N., Mascaro, O. and Chevallier, C. (2012). "Preschoolers Are Able to Take Merit into Account When Distributing Goods," *Developmental Psychology*, 48 (2), 492–498.

Bittles, A. and Neel, J. (1994). "The Costs of Human Inbreeding and Their Implications for Variation at the DNA Level," *Nature Genetics*, 8, 117–121.

Bliege Bird, R. and Power, E. (2015). "Prosocial Signaling and Cooperation Among Martu Hunters," *Evolution and Human Behavior*, 36 (5), 389–397.

Bloch, M. and Sperber, D. (2002). "Kinship and Evolved Psychological Dispositions," *Current Anthropology*, 43 (5), 723–748.

Boehm, C. (2001). *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*. Cambridge, MA: Harvard University Press.

———. (2012). *Moral Origins: The Evolution of Virtue, Altruism, and Shame*. New York: Basic Books.

Bowles, S. and Gintis, H. (2013). *A Cooperative Species: Human Reciprocity and Its Evolution*. Princeton: Princeton University Press.

Boyd, R., Gintis, H., Bowles, S. and Richerson, P. (2003). "The Evolution of Altruistic Punishment," *Proceedings of the National Academy of Sciences of the United States of America*, 100, 3531–3535.

Boyd, R. and Richerson, P. (2009). "Culture and the Evolution of Human Cooperation," *Philosophical Transactions of the Royal Society, London, Biological Sciences*, 364 (1533), 3281–3288.

Boyer, P. (2001). *Religion Explained: The Evolutionary Origins of Religious Thought*. New York: Basic Books.

———. (2018). *Minds Make Societies: How Cognition Explains the World Humans Create*. New Haven, CT: Yale University Press.

Boyer, P. and Petersen, M. (2011). "The Naturalness of (Many) Social Institutions," *Journal of Institutional Economics*, 8 (1), 1–25.

Bugental, D. B. (2000). "Acquisition of the Algorithms of Social Life: A Domain-Based Approach," *Psychological Bulletin*, 126, 187–219.

Burnstein, E., Crandall, C. and Kitayama, S. (1994). "Some Neo-Darwinian Decision Rules for Altruism: Weighting Cues for Inclusive Fitness as a Function of the Biological Importance of the Decision," *Journal of Personality and Social Psychology*, 67, 773–789.

Buss, D. (ed.). (2015). *Handbook of Evolutionary Psychology* (2nd ed., Vol. 1 and 2). Hoboken, NJ: John Wiley & Sons.

Buss, D., Larsen, R., Westen, D. and Semmelroth, J. (1992). "Sex Differences in Jealousy: Evolution, Physiology, and Psychology," *Psychological Science*, 3 (4), 251–255.

Cashdan, E. (1982). "Egalitarianism Among Hunters and Gatherers," *American Anthropologist*, 84, 116–120.

Chagnon, N. (1988). "Life Histories, Blood Revenge, and Warfare in a Tribal Population," *Science*, 239, 985–992.

———. (1992). *Yanomamö—The Last Days of Eden*. New York: Harcourt, Brace, Jovanovich.

Charlesworth, B. and Charlesworth, D. (1999). "The Genetic Basis of Inbreeding Depression," *Genetics Research*, 74, 329–340.

Choi, J. and Bowles, S. (2007). "The Coevolution of Parochial Altruism and War," *Science*, 318, 636–640.

Cosmides, L. (1985). "Deduction or Darwinian Algorithms? An Explanation of the 'Elusive' Content Effect on the Wason Selection Task," Doctoral dissertation, Department of Psychology, Harvard University, University Microfilms #86-02206.

———. (1989). "The Logic of Social Exchange: Has Natural Selection Shaped How Humans Reason? Studies with the Wason Selection Task," *Cognition*, 31, 187–276.

Cosmides, L. and Tooby, J. (1987). "From Evolution to Behavior: Evolutionary Psychology as the Missing Link," in J. Dupre (ed.), *The Latest on the Best: Essays on Evolution and Optimality*. Cambridge, MA: MIT Press.

———. (1989). "Evolutionary Psychology and the Generation of Culture, Part II. Case Study: A Computational Theory of Social Exchange," *Ethology & Sociobiology*, 10, 51–97.

———. (1992). "Cognitive Adaptations for Social Exchange," in J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press.

———. (1994). "Origins of Domain-Specificity: The Evolution of Functional Organization," in L. Hirschfeld and S. Gelman (eds.), *Mapping the Mind: Domain-Specificity in Cognition and Culture*. New York: Cambridge University Press.

———. (2006). "Evolutionary Psychology, Moral Heuristics, and the Law," in G. Gigerenzer and Christoph Engel (eds.), *Heuristics and the Law* (Dahlem Workshop Report 94). Cambridge, MA: MIT Press.

———. (2008a). "Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters," in W. Sinnott-Armstrong (ed.), *Moral Psychology*. Cambridge, MA: MIT Press, 53–119.

———. (2008b). "Can Evolutionary Psychology Assist Logicians? A Reply to Mallon," in W. Sinnott-Armstrong (ed.), *Moral Psychology*. Cambridge, MA: MIT Press, 131–136.

———. (2008c). "When Falsification Strikes: A Reply to Fodor," in W. Sinnott-Armstrong (ed.), *Moral Psychology*. Cambridge, MA: MIT Press, 143–164.

———. (2013). "Evolutionary Psychology: New Perspectives on Cognition and Motivation," *Annual Review of Psychology*, 64, 201–229.

———. (2015). "Adaptations for Reasoning About Social Exchange," in D. Buss (ed.), *The Handbook of Evolutionary Psychology, Second edition. Volume 2: Integrations*. Hoboken, NJ: John Wiley & Sons, 625–668.

Cosmides, L., Barrett, H. C. and Tooby, J. (2010). "Adaptive Specializations, Social Exchange, and the Evolution of Human Intelligence," *Proceedings of the National Academy of Sciences USA*, 107, 9007–9014.

Curry, O. (2015). "Morality as Cooperation: A Problem-Centred Approach," in T. Shackelford and R. Hansen (eds.), *The Evolution of Morality*. New York: Springer, 27–51.

Dawkins, R. (1979). "Twelve Misunderstandings of Kin Selection," *Ethology*, 51 (92), 184–200.

Debove, S., André, J-B. and Baumard, N. (2015). "Partner Choice Creates Fairness in Humans," *Proceedings of the Royal Society B*, 282, 392–399.

Debove, S., Baumard, N. and André, J-B. (2015). "Evolution of Equal Division Among Unequal Partners," *Evolution*, 69, 561–569.

Delton, A. W. and Cimino, A. (2010). "Exploring the Evolved Concept of Newcomer: Experimental Tests of a Cognitive Model," *Evolutionary Psychology*, 8 (2), 317–335.

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E. and Tooby, J. (2012). "The Psychosemantics of Free Riding: Dissecting the Architecture of a Moral Concept," *Journal of Personality and Social Psychology*, 102 (6), 1252–1270.

Delton, A. W, Krasnow, M. M., Cosmides, L. and Tooby, J. (2011). "Evolution of Direct Reciprocity Under Uncertainty Can Explain Human Generosity in One-Shot Encounters," *Proceedings of the National Academy of Sciences*, 108, 13335–13340.

Delton, A. W., Nemirow, J., Robertson, T. E., Cimino, A. and Cosmides, L. (2013). "Merely Opting Out of a Public Good Is Moralized: An Error Management Approach to Cooperation," *Journal of Personality and Social Psychology*, 105 (4), 621–638.

DeScioli, P. and Kurzban, K. (2013). "A Solution to the Mysteries of Morality," *Psychological Bulletin*, 139 (2), 477–496.

Eisenbruch, A., Grillot, R., Maestripieri, D. and Roney, J. (2016). "Evidence of Partner Choice Heuristics in a One-Shot Bargaining Game," *Evolution and Human Behavior*, 37, 429–439.

Everett, J., Pizarro, D. and Crockett, M. (2016). "Inference of Trustworthiness from Intuitive Moral Judgments," *Journal of Experimental Psychology: General*, 145 (6), 772–787.

Fehl, K., van der Post, D. and Semman, D. (2011). "Co-Evolution of Behaviour and Social Network Structure Promotes Cooperation," *Ecology Letters*, 14, 546–551.

Fehr, E. and Gachter, S. (2000). "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90, 980–994.

Fehrler, A. and Przepiorka, W. (2013). "Charitable Giving as a Signal of Trustworthiness: Disentangling the Signaling Benefits of Altruistic Acts," *Evolution and Human Behavior*, 34, 139–145.

Fessler, D. and Navarrete, C. (2004). "Third-Party Attitudes Toward Sibling Incest: Evidence for Westermarck's Hypotheses," *Evolution and Human Behavior*, 25, 277–294.

Fiddick, L., Cosmides, L. and Tooby, J. (2000). "No Interpretation Without Representation: The Role of Domain-Specific Representations and Inferences in the Wason Selection Task," *Cognition*, 77, 1–79.

Fiske, A. (1991). *Structures of Social Life: The Four Elementary Forms of Human Relationships: Communal Sharing, Authority Ranking, Equality Matching, Market Pricing*. New York: Free Press.

Fortes, M. (1970). *Kinship and the Social Order: The Legacy of Lewis Henry Morgan* (Morgan Lectures 1963). Oxford: Taylor & Francis.
Fox, R. (1965/1984) *The Red Lamp of Incest*. Notre Dame, IN: The University of Notre Dame Press.
Gigerenzer, G. and Hug, K. (1992). "Domain-Specific Reasoning: Social Contracts, Cheating, and Perspective Change," *Cognition*, 43 (2), 121–171.
Gill, M. and Nichols, S. (2008). "Sentimentalist Pluralism: Moral Psychology and Philosophical Ethics," *Philosophical Issues*, 18, 143–163.
Gintis, H. (2000). "Strong Reciprocity and Human Sociality," *Journal of Theoretical Biology*, 206, 169–179.
Gintis, H., Bowles, S., Boyd, R. and Fehr, E. (2003). "Explaining Altruistic Behavior in Humans," *Evolution and Human Behavior*, 24, 153–172.
Goldberg, S., Muir, R. and Kerr, J. (eds.). (2000). *Attachment Theory: Social, Developmental, and Clinical Perspectives*. London: The Analytic Press.
Greene, J. (2008). "The Secret Joke of Kant's Soul," in W. Sinnott-Armstrong (ed.), *Moral Psychology, Volume 3 Moral Psychology: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. Cambridge, MA: MIT Press, 35–79.
Güth, W. and Kocher, M. G. (2014). "More Than Thirty Years of Ultimatum Bargaining Experiments: Motives, Variations, and a Survey of the Recent Literature," *Journal of Economic Behavior & Organization*, 108, 396–409.
Haidt, J. (2012). *The Righteous Mind: Why Good People Are Divided by Politics and Religion*. New York: Vintage.
Hamann, K., Bender, J. and Tomasello, M. (2014). "Meritocratic Sharing Is Based on Collaboration in 3-Year-Olds," *Developmental Psychology*, 50 (1), 121–128.
Hamilton, W. (1964). "The Genetical Evolution of Social Behavior," *Journal of Theoretical Biology*, 7, 1–16.
Henrich, J. (2015). *The Secret of Our Success*. Princeton: Princeton University Press.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Smith Henrich, J., Hill, K., Gill-White, F., Gurven, M., Marlowe W. F., Patton Q. J. and Tracer, D. (2005). "Economic Man in Cross-Cultural Perspective: Behavioral Experiments in 15 Small-Scale Societies," *Behavioral and Brain Sciences*, 28 (6), 795–855.
Henrich, J., Boyd, R. and Richerson, P. (2012). "The Puzzle of Monogamous Marriage," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367 (1589), 657–669.
Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D. and Ziker, J. (2010). "Markets, Religion, Community Size, and the Evolution of Fairness and Punishment," *Science*, 327, 1480–1484.
Henrich, J., Heine, S. and Norenzayan, A. (2010). "The Weirdest People in the World?" *Behavioral and Brain Sciences*, 33 (2/3), 1–75.
Henrich, J. and Henrich, N. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. New York: Oxford University Press.
Higham, J. (2014). "How Does Honest Costly Signaling Work?" *Behavioral Ecology*, 25, 8–11.
Hoffman, E., McCabe, K. and Smith, V. (1998). "Behavioral Foundations of Reciprocity: Experimental Economics and Evolutionary Psychology," *Economic Inquiry*, 36, 335–352.
Huemer, M. (2005). *Ethical Intuitionism*. New York: Palgrave Macmillan.
Hursthouse, R. & Pettigrove, G. (2016). "Virtue Ethics," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/ethics-virtue/
Kameda, T., Takezawa, M., Tindale, R. and Smith, C. (2002). "Social Sharing and Risk Reduction: Exploring a Computational Algorithm for the Psychology of Windfall Gains," *Evolution and Human Behavior*, 23, 11–33.
Kanngiesser, P. and Warneken, F. (2012). "Young Children Consider Merit When Sharing Resources with Others," *PLoS One*, 7, e43979.

Kaplan, H. and Hill, K. (1985). "Food Sharing Among Ache Foragers: Tests of Explanatory Hypotheses," *Current Anthropology*, 26 (2), 223–246.

Kaplan, H., Schniter, E., Smith, V. and Wilson, B. (2012). "Risk and the Evolution of Human Exchange," *Proceedings of the Royal Society, B*, 279 (1740), 2930–2935.

Keeley, L. (1996). *War Before Civilization: The Myth of the Peaceful Savage*. New York: Oxford University Press.

Kelly, R. (1995). *The Foraging Spectrum: Diversity in Hunter-Gatherer Lifeways*. Washington, DC: Smithsonian Institution Press.

Kraft-Todd, G., Yoeli, E., Bhanot, S. and Rand, D. (2015). "Promoting Cooperation in the Field," *Current Opinion in Behavioral Sciences*, 3, 96–101.

Krasnow, M. M., Cosmides, L., Pedersen, E. and Tooby, J. (2012). "What Are Punishment and Reputation for?" *PLoS One*, 7 (9), e45662.

Krasnow, M. M., Delton, A. W., Cosmides, L. and Tooby, J. (2015). "Group Cooperation Without Group Selection: Modest Punishment Can Recruit Much Cooperation," *PLoS One*, 10 (4), e0124561.

Krasnow, M. M., Delton, A. W., Tooby, J. and Cosmides, L. (2013). "Meeting Now Suggests We Will Meet Again: Implications for Debates on the Evolution of Cooperation," *Nature Scientific Reports*, 3, 1747. doi:10.1038/srep01747.

Kurzban, R. (2012). *Why Everyone (Else) Is a Hypocrite: Evolution and the Modular Mind*. Princeton: Princeton University Press.

Kurzban, R., Dukes, A. and Weeden, J. (2010). "Sex, Drugs, and Moral Goals: Reproductive Strategies and Views About Recreational Drugs," *Proceedings of the Royal Society of London: Series B: Biological Sciences*, 277, 3501–3508.

Kurzban, R., Tooby, J. and Cosmides, L. (2001). "Can Race Be Erased? Coalitional Computation and Social Categorization," *Proceedings of the National Academy of Sciences USA*, 98 (26), 15387–15392.

Leimgruber, K., Shaw, A., Santos, L. and Olson, K. (2012). "Young Children Are More Generous When Others Are Aware of Their Actions," *PLoS One*, 7 (10), e48292.

Lewis, D., Al-Shawaf, L., Conroy-Beam, D., Asao, K. and Buss, D. (2017). "Evolutionary Psychology: A How-to Guide," *American Psychologist*, 72 (4), 353–373.

Lieberman, D. and Lobel, T. (2012). "Kinship on the Kibbutz: Coresidence Duration Predicts Altruism, Personal Sexual Aversions and Moral Attitudes Among Communally Reared Peers," *Evolution and Human Behavior*, 33, 26–34.

Lieberman, D. and Patrick, C. (2018). *Objection: Disgust, Morality and the Law*. New York: Oxford University Press.

Lieberman, D., Tooby, J. and Cosmides, L. (2003). "Does morality Have a Biological Basis? An Empirical Test of the Factors Governing Moral Sentiments Relating to Incest," *Proceedings of the Royal Society London (Biological Sciences)*, 270 (1517), 819–826.

———. (2007). "The Architecture of Human Kin detection," *Nature*, 445, 727–731.

Liénard, P., Chevallier, C., Mascaro, O., Kiurad, P. and Baumard, N. (2013). "Early Understanding of Merit in Turkana Children," *Journal of Cognition and Culture*, 13, 57–66.

Lim, J. (2010). "Welfare Tradeoff Ratios and Emotions: Psychological Foundations of Human Reciprocity," Doctoral dissertation, Department of Anthropology, University of California, Santa Barbara. *UMI* Number: 3505288.

List, J. (2007). "On the Interpretation of Giving in Dictator Games," *Journal of Political Economy*, 115 (3), 482–493.

Macfarlan, S., Walker, R., Flinn, M. and Chagnon, N. (2014). "Lethal Coalitionary Aggression and Long-Term Alliance Formation Among Yanomamö Men," *Proceedings of the National Academy of Sciences USA*, 111 (47), 16662–16669.

Mackie, D., Smith, E. and Ray, D. (2008). "Intergroup Emotions and Intergroup Relations," *Personality and Social Psychology Compass*, 2, 1866–1880.

Masclet, D., Noussair, C., Tucker, S. and Villeval, M. C. (2003). "Monetary and Nonmonetary Punishment in the Voluntary Contributions Mechanism," *American Economic Review*, 93, 366–380.

Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

McCloskey, D. (2006). *The Bourgeois Virtues*. Chicago: University of Chicago Press.
McDonald, M., Navarrete, C. and van Vugt, M. (2012). "Evolution and the Psychology of Intergroup Conflict: The Male Warrior Hypothesis," *Philosophical Transactions of the Royal Society, B*, 367, 670–679.
McElreath, R. and Boyd, R. (2006). *Mathematical Models of Social Evolution: A Guide for the Perplexed*. Chicago: University of Chicago Press.
McNamara, J., Barta, Z., Frohmage, L. and Houston, A. (2008). "The Coevolution of Choosiness and Cooperation," *Nature*, 451, 189–192.
Noë, R. and Hammerstein, P. (1994). "Biological Markets: Supply and Demand Determine the Effect of Partner Choice on Cooperation, Mutualism, and Mating," *Behavioral Ecology and Sociobiology*, 35, 1–11.
———. (1995). "Biological Markets," *Trends in Ecology & Evolution*, 10, 336–339.
Olson, M. (1965). *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.
Oxoby, R. and Spraggon, J. (2008). "Mine and Yours: Property Rights in Dictator Games," *Journal of Economic Behavior & Organization*, 65 (3–4), 703–713.
Panchanathan, K. and Boyd, R. (2003). "A Tale of Two Defectors: The Importance of Standing for Evolution of Indirect Reciprocity," *Journal of Theoretical Biology*, 224, 115–126.
Petersen, M. (2012). "Social Welfare as Small-Scale Help: Evolutionary Psychology and the Deservingness Heuristic," *American Journal of Political Science*, 56 (1), 1–16.
Pietraszewski, D., Cosmides, L. and Tooby, J. (2014). "The Content of Our Cooperation, Not the Color of Our Skin: Alliance Detection Regulates Categorization by Coalition and Race, but Not Sex," *PLoS One*, 9 (2), e88534.
Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined*. New York: Penguin Classics.
Price, M., Cosmides, L. and Tooby, J. (2002). "Punitive Sentiment as an Anti-Free Rider Psychological Device," *Evolution and Human Behavior*, 23, 203–231.
Rai, T. and Fiske, A. (2011). "Moral Psychology Is Relationship Regulation: Moral Motives for Unity, Hierarchy, Equality, and Proportionality," *Psychological Review*, 118 (1), 57–75.
Raihani, N. and Barclay, P. (2016). "Exploring the Trade-off Between Quality and Fairness in Human Partner Choice," *Royal Society Open Science*, 3, 160510–160516.
Rand, D., Arbesman, S. and Christakis, N. (2011). "Dynamic Social Networks Promote Cooperation in Experiments with Humans," *Proceedings of the National Academy of Sciences USA*, 108, 19193–19198.
Richerson, P. and Boyd, R. (2006). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.
Roney, J. and Simmons, Z. (2017). "Ovarian Hormone Fluctuations Predict Within-Cycle Shifts in Women's Food Intake," *Hormones & Behavior*, 90, 8–14.
Ross, W. D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
Runes, D. (1983). "Dictionary of Philosophy," *Philosophical Library*, 338.
Schino, G. and Aureli, F. (2017). "Reciprocity in Group Living Animals: Partner Control Versus Partner Choice," *Biological Reviews*, 92, 665–672.
Seemanová, E. (1971). "A Study of Children of Incestuous Matings," *Human Heredity*, 21 (2), 108–128.
Sell, A., Sznycer, D., Al-Shawaf, L., Lim, J., Krauss, A., Feldman, A., Rascanu, R., Sugiyama, L., Cosmides, L. and Tooby, J. (2017). "The Grammar of Anger: Mapping the Computational Architecture of a Recalibrational Emotion," *Cognition*, 168, 110–128.
Sell, A., Tooby, J. and Cosmides, L. (2009). "Formidability and the Logic of Human Anger," *Proceedings of the National Academy of Sciences*, 106 (35), 15073–15078.
Shaw, A., Montinari, N., Piovesan, M., Olson, K., Gino, F. and Norton, M. (2014). "Children Develop a Veil of Fairness," *Journal of Experimental Psychology: General*, 143, 363–375.
Shaw, A. and Olson, K. (2012). "Children Discard a Resource to Avoid Inequity," *Journal of Experimental Psychology: General*, 141 (2), 382–395.
Shepher, J. (1983). *Incest: A Biosocial View*. New York: Academic Press.

Sidanius, J. and Pratto, F. (1999). *Social Dominance: An Intergroup Theory of Hierarchy and Oppression*. New York: Cambridge University Press.

Singer, P. (2005). "Ethics and Intuitions," *Journal of Ethics*, 9 (3&4), 331–352.

Smith, E. and Winterhalder, B. (1992). *Evolutionary Ecology and Human Behavior*. New York: Walter de Gruyter.

Smith, V. (2003). "Constructivist and Ecological Rationality in Economics: Nobel Prize Lecture, December 8, 2002," Published in: *American Economic Review*, 93 (3), 465–508.

Stratton-Lake, P. (2016). "Intuitionism in Ethics," in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition). https://plato.stanford.edu/archives/win2016/entries/intuitionism-ethics/

Swanton, C. (2003). *Virtue Ethics: A Pluralistic View*. Oxford: Oxford University Press.

Sylwester, K. and Roberts, G. (2013). "Reputation-Based Partner Choice Is an Effective Alternative to Indirect Reciprocity in Solving Social Dilemmas," *Evolution and Human Behavior*, 34, 201–206.

Sznycer, D., De Smet, D., Billingsley, J. and Lieberman, D. (2016). "Coresidence Duration and Cues of Maternal Investment Regulate Sibling Altruism Across Cultures," *Journal of Personality and Social Psychology*, 111 (2), 159–177.

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S. and Halperin, E. (2016). "Shame Closely Tracks the Threat of Devaluation by Others, Even Across Cultures," *Proceedings of the National Academy of Sciences USA*, 113 (10), 2625–2630.

Tooby, J. (1982). "Pathogens, Polymorphism, and the Evolution of Sex," *Journal of Theoretical Biology*, 97, 557–576.

Tooby, J. and Cosmides, L. (1988). "The Evolution of War and Its Cognitive Foundations," Institute for Evolutionary Studies Technical Report #88–81.

———. (1990a). "The past explains the present: Emotional adaptations and the structure of ancestral environments," *Ethology and Sociobiology*, 11, 375–424. doi: 10.1016/0162-3095(90)90017-Z

———. (1990b). "On the Universality of Human Nature and the Uniqueness of the Individual: The Role of Genetics and Adaptation," *Journal of Personality*, 58, 17–67.

———. (1992). "The Psychological Foundations of Culture," in J. Barkow, L. Cosmides and J. Tooby (eds.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*. New York: Oxford University Press, 19–136.

———. (1996). "Friendship and the Banker's Paradox: Other Pathways to the Evolution of Adaptations for Altruism," in W. G. Runciman, J. Maynard Smith and R. I. M. Dunbar (eds.), *Evolution of Social Behaviour Patterns in Primates and Man: Proceedings of the British Academy*, 88, 119–143.

———. (2008). "The Evolutionary Psychology of the Emotions and Their Relationship to Internal Regulatory Variables," in M. Lewis, J. Haviland-Jones and L. Feldman Barrett (eds.), *Handbook of Emotions* (3rd ed.). New York: Guilford Press.

———. (2010). "Groups in Mind: Coalitional Psychology and the Roots of War and Morality," in Henrik Høgh-Olesen (ed.), *Human Morality and Sociality: Evolutionary and Comparative Perspectives*. London, UK: Palgrave Macmillan, 191–234.

Tooby, J., Cosmides, L. and Barrett, H. C. (2005). "Resolving the Debate on Innate Ideas: Learnability Constraints and the Evolved Interpenetration of Motivational and Conceptual Functions," in P. Carruthers, S. Laurence and S. Stich (eds.), *The Innate Mind: Structure and Content*. New York: Oxford University Press.

Tooby, J., Cosmides, L. and Price, M. (2006). "Cognitive Adaptations for n-Person Exchange: The Evolutionary Roots of Organizational Behavior," *Managerial and Decision Economics*, 27, 103–129.

Tooby, J., Cosmides, L., Sell, A., Lieberman, D. and Sznycer, D. (2008). "Internal Regulatory Variables and the Design of Human Motivation: A Computational and Evolutionary Approach," in Andrew J. Elliot (ed.), *Handbook of Approach and Avoidance Motivation*. Mahwah, NJ: Lawrence Erlbaum Associates, 251–271.

Trivers, R. (1971). "The Evolution of Reciprocal Altruism," *The Quarterly Review of Biology*, 46, 35–57.

———. (1974). "Parent-Offspring Conflict," *American Zoologist*, 14 (1), 249–264.

Tyber, J., Lieberman, L., Kurzban, R. and DeScioli, P. (2013). "Disgust: Evolved Function and Structure," *Psychological Review*, 120 (1), 65–84.

Van Vugt, M., De Cremer, D. and Janssen, D. (2007). "Gender Differences in Cooperation and Competition: The Male-Warrior Hypothesis," *Psychological Science*, 18, 19–23.
von Rueden, C. (2014). "The Roots and Fruits of Social Status in Small-Scale Human Societies," in J. Cheng, J. Tracy and C. Anderson (eds.), *The Psychology of Social Status*. New York: Springer, 179–200.
Westermarck, E. (1891/1921) *The History of Human Marriage* (5th ed.). London: Palgrave Macmillan.
Williams, G. C. (1966). *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton: Princeton University Press.
Williams, G. C. and Williams, D. (1957). "Natural Selection of Individually Harmful Social Adaptations Among Sibs with Special Reference to Social Insects," *Evolution*, 11, 32–39.
Wolfe, A. (1995). *Sexual Attraction and Childhood Association: A Chinese Brief for Edward Westermarck*. Redwood City, CA: Stanford University Press.
Wrangham, R. (In press 2019). *The Goodness Paradox: The Strange Relationship Between Virtue and Violence in Human Evolution*. New York, NY: Pantheon.
Wrangham, R. and Peterson, D. (1997). *Demonic Males: Apes and the Origins of Human Violence*. New York: Houghton-Mifflin.
Wrangham, R., Wilson, M. and Muller, M. (2006). "Comparative Rates of Violence in Chimpanzees and Humans," *Primates*, 47, 14–26.
Yamagishi, T. (1986). "The Provision of a Sanctioning System as a Public Good," *Journal of Personality and Social Psychology*, 51, 110–116.

Further Readings

For conceptual foundations of evolutionary psychology see John Tooby and Leda Cosmides "The Psychological Foundations of Culture," in Jerome Barkow, Leda Cosmides and John Tooby, eds.), *The Adapted Mind* (New York: Oxford University Press, 1992) and Part I, Volume 1 of David Buss, ed., *The Handbook of Evolutionary Psychology* (2nd ed.) (Hoboken, NJ: John Wiley & Sons, 2016). Volumes 1 and 2 of this handbook present current research on many topics, including ones relevant to moral epistemology. For how fairness in cooperation can evolve via partner choice, see Nicolas Baumard, Jean-Baptiste André and Dan Sperber, "A Mutualistic Approach to Morality: The Evolution of Fairness by Partner Choice," *Behavioral & Brain Sciences*, 36, 59–78, 2013. For an examination of evidence (including counterhypotheses) for a reasoning system specialized for social exchange and detecting cheaters, see Leda Cosmides and John Tooby, "Can a General Deontic Logic Capture the Facts of Human Moral Reasoning? How the Mind Interprets Social Exchange Rules and Detects Cheaters," in Walter Sinnott-Armstrong (ed.), *Moral Psychology* (Cambridge, MA: MIT Press, 2008). Jonathan Haidt (2012) discusses multiple moral domains and the role of evolved moral intuitions versus reasoning in moral judgment in *The Righteous Mind: Why Good People Are Divided by Politics and Religion* (New York: Pantheon). For the link between disgust and morality, see Debra Lieberman and Carlton Patrick's *Objection: Disgust, Morality and the Law* (New York: Oxford University Press, 2018). On moral norms and impartiality arising from coalitional psychology and alliance formation, see John Tooby and Leda Cosmides, "Groups in Mind: Coalitional Psychology and the Roots of War and Morality," in Henrik Høgh-Olesen (ed.), *Human Morality and Sociality: Evolutionary and Comparative Perspectives* (London, UK: Palgrave Macmillan, 2010). For how evolutionary psychology relates to cultural transmission, including morality, religion, and social institutions, see Pascal Boyer's *Minds Make Societies: How Cognition Explains the World Humans Create* (New Haven, CT: Yale University Press, 2018).

Related Chapters

Chapter 1 The Quest for the Boundaries of Morality; Chapter 2 The Normative Sense: What is Universal? What Varies? Chapter 3 Normative Practices of Other Animals; Chapter 4 The Neurological Basis of Moral Psychology; Chapter 5 Moral Development in

Humans, Chapter 6 Moral Learning; Chapter 7 Moral Reasoning and Emotion; Chapter 8 Moral Intuitions and Heuristics; Chapter 12 Contemporary Moral Epistemology; Chapter 13 The Denial of Moral Knowledge; Chapter 14 Nihilism and the Epistemic Profile of Moral Judgment; Chapter 15 Relativism and Pluralism in Moral Epistemology; Chapter 16 Rationalism and Intuitionism: Assessing Three Views about the Psychology of Moral Judgment; Chapter 18 Moral Intuition; Chapter 20 Moral Theory and its Role in Everyday Moral Thought and Action; Chapter 22 Moral Knowledge as Know-How; Chapter 23 Group Moral Knowledge; Chapter 28 Decision Making Under Moral-Uncertainty; Chapter 29 Public Policy and Philosophical Accounts of Desert.