

Culture, evolution and the puzzle of human cooperation

Action editors: Luca Tummolini and Cristiano Castelfranchi

Joseph Henrich *, Natalie Henrich

Department of Anthropology, Emory University, Geosciences Building, 1557 Pierce Drive, Atlanta, GA 30322-1720, United States

Received 15 March 2005; accepted 7 November 2005

Available online 9 March 2006

Abstract

Synthesizing existing work from diverse disciplines, this paper introduces a culture-gene coevolutionary approach to human behavior and psychology, and applies it to the evolution of cooperation. After a general discussion of cooperation in humans, this paper summarizes Dual Inheritance Theory and shows how cultural transmission can be brought under the Darwinian umbrella in order to analyze how culture and genes coevolve and jointly influence human behavior and psychology. We then present a generally applicable mathematical characterization of the problem of cooperation. From a Dual Inheritance perspective, we review and discuss work on kinship, reciprocity, reputation, social norms, and ethnicity, and their application to solving the problem of cooperation.

© 2006 Published by Elsevier B.V.

Keywords: Cooperation; Culture-gene coevolution; Reciprocity; Reputation; Punishment; Norms; Ethnicity; Cultural transmission; Dual inheritance theory

It must not be forgotten that although a high standard of morality gives but a slight or no advantage to each individual man and his children over the other men of the same tribe, yet that an increase in the number of well-endowed men and an advancement in the standard of morality will certainly give an immense advantage to one tribe over another. A tribe including many members who, from possessing in a high degree the spirit of patriotism, fidelity, obedience, courage, and sympathy, were always ready to aid one another, and to sacrifice themselves for the common good, would be victorious over most other tribes; and this would be natural selection. At all times throughout the world tribes have supplanted other tribes; and as morality is one important element in their success, the standard of morality and the number of well-endowed men will thus everywhere tend to rise and increase.

–Charles Darwin (*Descent of Man* (1871, Chapter 5)).

The tendency to imitate may come into direct conflict with the prudential teachings of pleasure and pain, and yet may be acted upon. A child may do, and keep on doing, imitations which cause him pain.

–James Mark Baldwin (*Mental Development in the Child and the Race* (1906, Chapter 10)).

1. The why puzzle¹

All around us we see people contributing to the welfare of others, even when it is not convenient and may be costly in terms of time or money, or may affect their personal and professional relationships. In fact, we see so much of this cooperation in daily life that we usually don't notice it or stop to question *why* people are bothering to help others, or how such seemingly ubiquitous

* Corresponding author. Tel.: +1 404 727 5248; fax: +1 404 727 2860.
E-mail address: jhenric@emory.edu (J. Henrich).

¹ The paper synthesizes material drawn from our book, *Why People Cooperate* (in press).

cooperation might be explained. When asked why someone cooperates, a common reply is that helping others is “the right thing to do” and that people “should” help each other. Some scholars have merely accepted such cooperative acts as part of being human (Durkheim, 1933), without endeavoring to question why we cooperate sometimes, but not others, or why different societies seem to cooperate to differing degrees. In fact, not only are there times that we don’t cooperate when we know we could have, there are many times when we don’t even perceive an opportunity for cooperation when one exists. Taking this commonsense observation as a point of departure, we will address the question of *when and why people incur personal costs in order to help another person or group of people*.

Even casual observation suggests some robust patterns in cooperative social behavior. You’ve probably noticed that people can be quite particular about who they will help, when, and how much. First, people act frequently, and sometimes at great cost to themselves, to help their families, especially their kids. Why is that? People help friends, and sometimes acquaintances, but there is something different about ‘the rules’ for helping these people vs. helping close family. Friends who break the ‘helping rules’ often drop from ‘friend’ to ‘acquaintance’. Kids, on the other hand (even as adults), are not only given more latitude, but are evaluated using quite different ‘rules’. Why are friends different from family members, and where do these ‘rules’ come from? Moreover, what about helping strangers? For instance, have you ever considered why you would willingly stop to give directions to a lost visitor on the street rather than continuing onwards? Such an action wastes your time, has some risks (the person may be a thief, con-artist or murderer), and you’ll likely never see that person again. If you don’t stop to help this stranger, why would you feel bad about it? Are you more likely to help some strangers than others? If so, who are you more likely to help, and why?

Addressing such *why*-questions requires distinguishing at least two levels of explanation. The first, the *proximate psychological level*, focuses on understanding the psychological processes and preferences that propel certain decisions and behaviors. For example, how does the psychology ‘loving your kin’ work? Who qualifies as ‘kin’? Are there different kinds of ‘kin’ who get different amounts of love or help? The second level, the *ultimate evolutionary level*, explores the evolutionary processes that produced the proximate psychologies that, in turn, produce decisions and behaviors. If you are already familiar with evolutionary theories that seek to explain the kinds of social behavior mentioned above, you may still find the following discussion interesting, as in our view ‘evolutionary theory’ includes both cultural evolution and culture-gene coevolution. The framework below integrates what we call the ‘canonical models’ of cooperation (e.g., kin selection and reciprocal altruism)

within a broader coevolutionary or dual inheritance framework.²

This paper lays out a set of evolutionary theories, derived from a single framework, aimed at explaining the ultimate origins of different aspects of cooperative behavior. Applying these ultimate theories, we are able to derive predictions and set up research questions about the details of the proximate psychological mechanisms for social behavior and cooperation. Thus, dealing with the ultimate *why*-question allows us to better address the proximate psychological (and behavioral) questions. Simultaneously, dealing with the proximate psychological issues provides the only way to fully test our ideas about ultimate causes.

2. What do we mean by ‘cooperation’ and ‘prosociality’?

Cooperation occurs when an individual incurs a cost in order to provide a benefit for another person or people. Costs include things that relate to genetic fitness like resources (e.g., money, time, labor, and food). Throughout our discussions we often refer to cooperative acts as ‘giving help’ – but cooperative acts are not limited to giving help. Cooperative interactions take place within pairs, small groups or large groups, and can occur among friends, relatives or strangers. In pairs, cooperation might involve babysitting, giving a friend a ride to the airport, loaning sugar to a neighbor, or making dinner for a sick person. Among large groups, examples of cooperation include voting, participating in Neighborhood Watch, recycling, contributing to public radio, sharing food, and paying taxes. In these cases, a large group of people benefit from the costly actions of an individual. To fully understand why these behaviors qualify as cooperation, let’s look more closely at three of them: voting, food sharing and recycling.

When a person goes to the voting booth on Election Day, she incurs a cost. There is the time it takes to drive or walk to the voting place, the time to vote, and the time to return back to work (in the US, unlike many countries, voting occurs on a work day instead of a holiday). There may be a financial cost if the voter has to pay for parking and/or gas, and there is the opportunity cost in that the person could have been using this time to do something else, like finishing a report for the boss or spending time with her children. Benefits also occur from voting, namely

² The coevolutionary approach championed here stems directly from the research program laid out by Darwin in the *Descent of Man* (1871), and subsequently developed by Baldwin (Baldwin, 1896a, 1896b). At the opening of this paragraph, we wrote “at least two levels of explanation” because with the emergence of cultural capacities and cultural evolution, it is sometimes useful to distinguish between levels of explanation: (1) *Ultimate level*. Natural selection builds the psychological capacities for cultural learning; (2) *Intermediate level*. Culture evolves, accumulates and adapts non-genetically (intermediate level) to produce local skills, preferences, and cognitive abilities; (3) *Proximate level*. Psychological mechanisms that are the joint product of genetic and cultural evolutionary history propel decision-making and behavior.

the support of the democratic system and the electoral process – if no one voted, the system would collapse and the government could be deemed illegitimate. All members of the voting community share this benefit. Contrary to what many voters believe, however, one of the benefits of voting is *not* that their candidate of choice is more likely to be elected because of the ballot that the individual voter cast. Have you ever heard of a president being voted into office by a margin of one vote? While it is true that the democratic system would collapse if no one voted, it is also true that any single vote has a negligible effect. Therefore, when a person votes she performs a costly act that helps the group (i.e., preserves democracy) but does not help herself (since her candidate will not be elected because of her vote), and she incurs the costs associated with voting. Consequently, voting qualifies as a large-group cooperative act.

Food sharing, also a cooperative act, can be found throughout the world. Within our own society, food sharing includes offering some of your meal to others when eating at a restaurant, providing food and drink to guests in your home, and donating non-perishables to food banks. In hunting-and-gathering societies food sharing is wide spread and important. Among groups such as the Aché of Paraguay (Hill, 2002) and the Hadza of Tanzania (Hawkes, O’Connell, & Blurton Jones, 2001), meat sharing is the norm, with hunters routinely sharing their kill with the group. In sharing, the hunter incurs a personal cost (his time, effort and the loss of some of the meat that he and/or his family could have consumed alone) while benefiting the rest of the group by providing them with meat. Like voting, the distribution of costs and benefits qualifies food sharing as a form of cooperative behavior, which likely has deep evolutionary importance.

Recycling is an excellent example of large-scale cooperation. In the 1990s, vigorous campaigns were launched to promote recycling of paper, glass, metal and plastic products. In many North American cities, each household received boxes for each type of recyclable material and the city provided regular pick-up of these items. When we recycle we incur several costs, including washing out cans and bottles rather than tossing the dirty containers directly into the garbage can, sorting our garbage by product type, and taking a multitude of containers to the curb on collection day (assuming we can even remember which day each type of recyclable is collected!). Although this process is not a huge burden, it certainly takes more time and effort than simply throwing all our refuse into the trash, and often requires special containers, more attention and memory, and more household space. Now consider the benefits from these individual-level costs: the planet and all of its inhabitants get to live in a cleaner, healthier world. What happens if you (one person) decide not to recycle? Nothing happens. The contribution of any one person is insignificant in terms of the planet’s health, and you and your kids still get to live in a cleaner world, as long as many

other people pay the costs of recycling.³ The benefits that you personally create by recycling do not outweigh your costs, and consequently any self-interested, rational individual who weighed the costs and benefits of his actions should refrain from recycling. Yet, if everyone did this, the environment would suffer.

Recycling, like voting, is an example of a public goods problem. In a public goods problem any given individual has an incentive to refrain from cooperating even though in the long run everyone will suffer from the loss of the joint benefit. In public goods situations individuals can free-ride by not cooperating while still reaping the benefits created by the contributions of others. If you have ever been in a group project in school, in which the group receives a single grade for its efforts, you may have experienced free-riding. In this context, free-riders avoid doing work but still reap the benefits (the grade) of the others’ work. In the recycling example, a free-rider would be someone who chooses not to recycle, but still enjoys breathing the cleaner air that results from those who do.

Besides cooperative behaviors in which an individual provides a direct benefit to others at a cost to herself, there is a larger class of behaviors that we will call prosocial. The best example of a prosocial behavior is what economists call altruistic punishment (Fehr & Gächter, 2002). Here, an individual pays a cost to inflict a cost on another individual in order to maintain an individually costly behavior in a group. For example, in the gasoline crisis of the 1970s there were long lines at the pumps. Occasionally, individuals would attempt to free-ride by entering the line near the front. Inevitably, this free-riding would infuriate at least one person in the line who would often threaten, and if necessary physically assault, the ‘line-jumper’. The presence of the punishers no doubt dissuaded some individuals from entering the line near the front. This altruistic punishment is costly for the punishers, as he or she (it was usually a he) risked getting beat-up, but benefits the group by maintaining orderly lines.⁴ Non-punishers in the line reap the benefits created by punishers without having to pay the cost of wrestling with frustrated and enraged line-jumpers. As we will discuss more below, theoretical models have repeatedly shown that if individuals are willing to punish others at a cost to themselves, some otherwise puzzling forms of cooperation can be explained. If there are altruistic punishers out there, anyone who can learn will fall into line, even if he is completely selfish. Furthermore, experimental evidence clearly shows that, at least in some societies, people are willing to punish anonymous strangers at a cost to

³ We realize that there is some debate about whether recycling, when all the costs and effects are taken in account, will actually achieve these beneficial ends. However, as will be come clear below, social norms can potentially maintain any behavior, even ones that don’t yield overall benefits.

⁴ Those who doubt the group benefit created by orderly lines have not spent much time in countries where it is not the custom to make orderly lines for scarce resources.

themselves (Henrich et al., 2004). Thus, our term prosocial encompasses both cooperation as we described above ('helping'), and altruistic punishment. We have avoided extending 'cooperation' to cover altruistic punishment because of the lack of fit with common intuitive understandings of 'cooperation'.

3. Why is it so hard to explain cooperation and prosociality?

As we saw above, cooperation and altruistic punishment always involve a cost to the cooperator or punisher. This led many evolutionary and rational choice scholars to ask the natural question: if cooperation is costly to the individual, why does anyone do it? Evolutionary biologists like Dawkins (1976) emphasize the logic through which natural selection produces "selfish genes". The logic suggests that genes that, on-average, cause their bearers (the individuals) to pay fewer costs and reap more benefits relative to others, are the ones more likely to be transmitted into the next generation.⁵ If a person has a gene that leads her to incur costs to help other individuals, (i.e., provides benefits to these people), then this individual will, on-average, produce fewer offspring than individuals who do not possess these genes. Taken at face value, this verbal reasoning indicates that cooperation will generally be selected against by natural selection, and that cooperation ought to be rare, both in humans and throughout nature.⁶

Let's illustrate this with a more concrete, and ethnographically relevant, example. Consider a person with a gene that leads her to engage in sharing food.⁷ By sharing her food with the group, she is increasing the fitness of everyone else in the group by providing them with extra calories and nutrients. However, at the same time she is lowering her fitness and the fitness of her offspring by taking away food from them and giving it to others in the group. All else being equal, her generosity will result in her rearing fewer healthy offspring to adulthood than a person who avoids sharing. The generous food sharer will likely have some daughters who also share food with the group, and these daughters, like their mother, will have fewer offspring than the non-sharers in their group. As you can see, in each generation, the frequency of the 'sharing gene' will decrease because of the behavior it promotes. Even if 'food sharing genes' were initially very common, they would gradually disappear from a population over

many generations. Nevertheless, we see lots of food sharing in the world.

While the broad thrust of theoretical evolutionary biology suggests that cooperation and prosociality should generally be rare in nature, there has arisen in the last 40 years a vibrant and growing set of theoretical models (theoretical = mathematical) that have demonstrated a variety of evolutionary pathways to cooperation. These pathways are not mutually exclusive solutions to the dilemma of cooperation, and different pathways may cross in ways that either facilitate more cooperation or debilitate it by creating conflicts. Any particular organism may have evolved to make use of one, two, all of them, or none of them. Some of the pathways or 'classes of models' that we discuss below are applicable to an enormous range of species, while others are premised on a heavy reliance on high fidelity cultural learning, and thus may be restricted to humans. The second half of this paper provides an introduction to each of the models, and begins to track down their empirical entailments. Before we turn to the models, we first need to layout some background information on social learning and evolutionary psychology, as these concepts underlie important components of subsequent discussions.

3.1. The puzzle deepens for our species

The puzzle of cooperation is both more interesting and more enigmatic for our species. This is because the nature of human cooperation, while similar in some ways to the patterns observed in other species, is quite different in several key respects. At a macro level, human cooperation varies substantially from non-human primates in both its scale and the nature of its variability. While the scale of cooperation in other primates rarely exceeds two or three individuals (e.g., in grooming and coalitions), humans in some societies, including many hunter-gatherer societies, cooperate on scales involving hundreds, thousands or even more (e.g., war, voting, recycling, and exchange networks). However, it is not merely the scale of cooperation that is different from all non-human mammalian species, but also the degree of variability across social groups. The scale of human cooperation varies dramatically across social groups, from societies that are economically independent at the family level – showing little cooperation outside the extended kin circle (e.g., Johnson & Earle, 2000: Machiguenga, Shoshone) – to the vast scales found in chiefdoms and modern states (Richerson & Boyd, 2000). While ecological factors are certainly part of the explanation for this variation, substantial degrees of variation in the scale of cooperation can be observed among social groups inhabiting identical environments (e.g., Atran et al., 1999; Itza Ladinós and Kekchi; Kelly, 1985: the Nuer and Dinka). Moreover, historical sources show that the scale of cooperation in many societies has increased by orders of magnitude in historical time (Diamond, 1997), thereby indicating the presence of some non-genetic evolutionary process that has been ratcheting up the scale of cooperation

⁵ This assumes that these 'costs' and 'benefits' translate in some manner (however weakly) into differences in survival and reproduction.

⁶ Economics has also sought to explain the puzzle of cooperation. Interestingly, however, the reasons why the puzzle has arisen in Economics has more to do with the disciplinary tradition of assuming individuals are self-interested rather than any *prima-facie* deductive logic. Other than its heritage in Enlightenment philosophy, there is no reason (that we know of) why Economics has typically assumed pure self-interest. Smith (2000), for example, wrote eloquently about the importance of moral sentiments.

⁷ The assumption of a 'single gene' is a rhetorical and modeling convenience and does not substantially impact the outcomes of these evolutionary arguments.

(a process that has not been observed in other species).⁸ Finally, while primate species typically show little variation in the behavioral domains of cooperative behavior, human social groups vary substantially in their domains of cooperation. For example, some groups cooperate in fishing, but not house-building or warfare, while other groups cooperate in house-building and warfare, but not fishing. A complete approach to the puzzle of human cooperation needs to be able to explain these patterns in a manner that links humans to the rest of the natural world, but at the same time explains our distinctiveness (Henrich, 2004; Henrich et al., 2003).

3.1.1. *Intelligence is not the answer*

Because the scale of human cooperation is so much greater than that found among other mammals, and particularly other primates, there is a common intuition that human cooperation must result from our ‘superior intelligence’. We believe this intuition is likely wrong for both theoretical and empirical reasons. First, a substantial amount of theoretical work in Economics, Anthropology and Biology shows that more intelligence usually leads to *less* cooperation, not more. Cognitive capacities for strategic thinking that include planning for the future, storing data on past interactions and more accurately assessing potential costs and benefits do not lead to more cooperation, as many people think (more on this below in the reciprocity subsection). In a complicated world with imperfect information, the skills of deception, deceit, trickery and manipulation, which are improved by some kinds of intelligence, are more powerful at destroying cooperation than are capacities for tracking past interactions and preserving it. There are always many easier ways to break something fragile than to protect it.

Second, empirically the ‘intelligence hypothesis’ does not lead to the kind of cooperation that characterizes our species. As just noted above, the scales of human cooperation vary dramatically across social groups and domains (even when groups inhabit the identical environment), and have changed over historical time. It is difficult to see how the ‘intelligence hypothesis’ could explain these fundamental patterns. All humans are smart, but some human societies actually cooperate less, and live in smaller groups than non-humans, while other human societies cooperate on massive scales. Moreover, scales of human cooperation have increased dramatically over the last 5000 years in many societies, but we have little reason to believe this dramatic increase resulted from genetic changes influencing human ‘intelligence’.

Finally, we get insight on the relationship between ‘intelligence’ and ‘cooperation’ by looking comparatively at other cooperative species. Besides humans, the next best cooperators in the animal kingdom are the eusocial insects

(bees, wasps, ants, etc.). These critters manage to achieve massive levels of cooperation – only achieved by humans in historical time – with very few neurons per individual. There are many species that have substantially *more* neurons (by several orders of magnitudes) than eusocial insects – including all primates – but all of these species cooperate *less* than these insects. There is apparently no necessary relationship in nature between intelligence and cooperation. Thus, other than a strong intuition, there is little to support the ‘intelligence hypotheses’.

4. Culture is a part of evolutionary theory and human biology

Since the rise of human sociobiology in the 1970s, *culture* and *biology* or *cultural explanations* and *evolutionary explanations* have often been opposed, and the seeming opposition between the categories has led to a great deal of unnecessary dispute and debate. This dichotomy, and the associated arguments, are outmoded and unproductive. A wide range of human behaviors, which most would think of as purely cultural (dress, greetings, food taboos, etc.), are actually 100% cultural and 100% genetic. Behaviors are *cultural* in that they are socially learned by observation and interaction in a social group. All culturally acquired behaviors, beliefs, preferences, strategies, practices (hereafter, we refer to all these collectively as ‘traits’) are also *genetic* in the sense that their acquisition requires brain machinery that allows for substantial amounts of complex, high-fidelity social learning. We know that there are ‘human genes’ that allow for cultural behavior, as chimpanzees raised (enculturated) alongside human children do not acquire anything approaching adult human behavioral patterns or social norms.⁹ In general, our species’ social learning capacities far outstrip all others (we are a hyper-cultural species), and this capacity can best be understood as a genetically-evolved adaptation for acquiring adaptive traits in complex, variable environments (Boyd & Richerson, 1985). For example, if you were born into a band of hunter-gatherers, it would likely be more adaptive for you to simply copy how the other members of your group make bows and arrows than to individually re-invent all the complex details of the manufacturing process (Henrich, *in press*). If you can take advantage of the experience of other hunters, and possibly the wisdom of previous generations that has been passed down through traditional practices, then natural selection will favor social learning, a.k.a. cultural capacities. With regard to cooperation, it doesn’t take much ethnographic reading to realize that understanding social learning is central to figuring out how, when, with whom and how much people cooperate. Because of its relevance to cooperation, we briefly explore some of the

⁸ We can rule out genetic evolution for these changes in social behavior because the time periods of these historical changes are too short.

⁹ Interestingly, one experiment of this kind was discontinued because, while the chimpanzees had ceased learning much human behavior from his adopted family, the chimp’s human brother had started acquiring many chimp-behaviors via imitation.

evolutionary theory underlying social learning before continuing with theory more directly linked to cooperation. In doing this, we put culture under the umbrella of Darwinian Theory.

5. The evolution of cultural capacities and cultural evolution

Dual Inheritance Theory allows culture to be fully incorporated into evolutionary theory. The approach can be summarized with three key ideas:

- (1) Culture, cultural transmission and cultural evolution arise from genetically evolved psychological adaptations for acquiring ideas, beliefs, values, practices, mental models, and strategies from other individuals by observation and imitation. Below, we summarize how evolutionary theory has been used to predict the psychological details of these cultural learning cognitive capacities.
- (2) These psychological mechanisms for social learning led to behaviors that were, *on-average*, adaptive in the varying ancestral environments that characterized our human lineage. Any particular individual's behavior or group's cultural practice may be adaptively neutral or maladaptive. By specifying some of the psychological details of these cultural learning abilities (see #1 above), cultural evolutionary models enable us to predict the patterns and conditions of maladaptation, and thus provide theories of both adaptation and of maladaptation (Boyd & Richerson, 1985, chap. 7). This is an advantage over the models traditionally used in sociobiology and human behavioral ecology, in which a behavior is either 'adaptive', or inexplicable.
- (3) The emergence of cultural learning capacities in the human lineage creates population processes that change the selective environments faced by genes. For example, suppose the practice of cooking meat spread by imitative learning in ancestral human populations. In an environment of 'cooked meat', natural selection may favor genes that shorten our energetically costly intestines and alter our digestive chemistry. Such a reduction of digestive tissue may have freed up energy for more 'brain building'. In this way, human biology is adapting to culturally transmitted behavior. The interactive effect is called culture-gene coevolution. As discussed below, this interaction may be critical for understanding some aspects of human cooperation, particularly large-scale cooperation among non-relatives (Baldwin, 1896a; Boyd & Richerson, 2002b; Durham, 1991; Henrich, 2004; Richerson & Boyd, 1998, 2000).

Below we expand on each of these ideas in greater detail, although for a complete understanding of this approach to cultural evolution, readers should begin with Henrich and McElreath (2003) or Richerson and Boyd (2005).

5.1. Evolved psychological mechanisms for learning culture

The approach of understanding culture using evolutionary theory begins by considering what kinds of cognitive learning abilities would have allowed individuals, in the changing environments of our hunter-gatherer ancestors, to efficiently and effectively extract adaptive ideas, beliefs, and practices from their social worlds. This approach diverges from mainstream evolutionary psychology in its emphasis on the *costly information hypothesis* and the evolution of specialized *social learning mechanisms*. The costly information hypothesis focuses on the evolutionary tradeoffs between acquiring accurate behavioral information at high cost (and less flexibility) and gleaning less accurate information at low cost (and greater flexibility). By formally exploring how the costly information hypothesis generates trade-offs in the evolution of our social learning capacities, we can generate predictive theories about the details of human cultural psychology (Henrich & McElreath, 2003). When information is costly, natural selection will favor cultural learning mechanisms that allow individuals to extract adaptive information, strategies, practices, heuristics and beliefs from other members of their social group at a lower cost than through alternative individual mechanisms (like trial-and-error learning). Human cognition probably contains numerous heuristics, directed attentional mechanisms and biased inferential tendencies that facilitate the acquisition of useful traits.

Such cultural learning mechanisms can be categorized into (1) *content biases* and (2) *context biases*. Content biases, or what Boyd and Richerson (1985) have called *direct biases*, cause us to more readily acquire certain beliefs, ideas or behaviors *because* some aspect of their content makes them more appealing. For example, imagine three practices involving different additives to popcorn: the first involves putting salt on popcorn, the second favors adding sugar, and the third involves sprinkling sawdust on the kernels. Innate content biases that affect cultural transmission will guarantee that sawdust will not be a popular popcorn additive in any human societies. An innovative company may try to market sawdust popcorn, but it is unlikely to spread as long as salt and sugar are out there as alternatives. Both salt and sugar have innate content biases for sensible evolutionary reasons.¹⁰ Of course, if you grew up in a society that only salts its popcorn, you may steadfastly adhere to your salting preference even once you find that sugar is the standard popcorn additive in many places. Many such innate biases may have evolved because they facilitate the acquisition of fitness-enhancing cultural traits. However, content biases may also be a byproduct (not specifically selected for) of other psychological processes,

¹⁰ Foods with salty or sugary flavors were likely both important sources of scarce nutrients and calories in ancestral human environments, and in short supply. Thus, natural selection should favor a bias to acquire a taste for salty or sweet foods.

or the product of cultural transmission.¹¹ Because content biases are likely numerous and generally confined to particular domains of culture, we omit further discussion of them here.

Context biases, on the other hand, guide social learning by exploiting cues from the individuals who are being learned from (we will term these individual ‘models’), rather than features of the thing being learned (the cultural trait). There is a great deal of adaptive information embodied in both *who* holds ideas and the *commonness* of the ideas or practices (i.e., the number or percent of people who have the trait). For example, because information is costly to acquire, individuals will do better if they preferentially pay attention to, and learn from, people who are highly successful, particularly skilled and well-respected. In fact, this is true even if the person’s skill or success is not directly connected to the behavior, belief or practice in question. A large amount of mathematical modeling effort has been expended in exploring the conditions under which different context biases will evolve, how they should be constructed psychologically, and what population patterns will emerge from individuals using such learning mechanisms. Moreover, and perhaps more importantly, a vast amount of field and laboratory data confirms that these learning biases are indeed an important part of our cognition. Our remaining discussion of psychological mechanisms focuses on two categories of context biases in cultural learning: (1) success and prestige bias and (2) conformity bias.

5.1.1. *Success and prestige bias*

If solutions to the problems of survival are tough to figure out on your own, but you can imitate others, who should you imitate? If possible, you should imitate people who are winning at the game of life – that is, successful people. More precisely, if individuals vary in skills (e.g., tool making), strategies (e.g., tracking techniques), and/or preferences (e.g., for foods) in ways that affect fitness, and at least some components of those differences can be acquired via cultural learning, then natural selection will favor cognitive capacities that cause individuals to preferentially learn from more successful individuals. The greater the variation in acquirable skills among individuals, and

the more difficult those skills are to acquire by individual learning, the greater the pressure to preferentially focus one’s attention on, and imitate, the most skilled individuals. In general, this prediction has been largely confirmed with data from across the social sciences: people are powerfully motivated to (unconsciously) imitate skilled and successful people (Henrich & Gil-White, 2001).

The problem of figuring out what things about a person to imitate is trickier than it might appear at first glance. Knowing that an individual is more successful than others does not tell the learner which of an individual’s many traits are responsible for the success. Is a businesswoman successful because of her interactional style, religious beliefs, professional attire, workout regime, vitamin preferences, or work ethic? Her success could be attributable to any, most, or all of these. Because of this ambiguity, the theory suggests that humans have likely evolved a propensity to copy a wide range of cultural traits from successful individuals, only some of which may actually relate to the individuals’ success (Henrich & Gil-White, 2001). One outcome of such a copying bias is that many neutral and some slightly maladaptive traits can hitch-hike along with adaptive cultural traits. For example, many successful business people, among other things, work long hours at the office and carry a leather briefcase. A person who wants to be successful at work may copy both of these practices, as well as many others. Working long hours is probably related to the model’s success whereas carrying a leather briefcase may be a neutral trait. Through this imitative process, generations of business people carry leather briefcases even though it doesn’t increase their success at work. In a world of costly information, cognitive adaptations don’t always produce adaptive behaviors from the point of view of genes and the theory allows for predictions about the conditions under which neutral and maladaptive cultural traits will spread.

The predictions derived from this approach have already been confirmed by a variety of evidence from across the social sciences. Adults and children do preferentially imitate (often unconsciously) successful and skilled individuals, and they imitate them in a variety of ways, even in areas well outside of the person’s domain of success. Henrich and Gil-White (2001) summarize much of the general evidence in this regard, and experimental economists have clearly demonstrated that people readily use imitation to figure out how to behave in social interactions involving cooperation and competition, especially when payoffs are on the line (Selten & Apesteegula, 2002).

5.1.2. *Conformist transmission*

What do you do when observable differences in behavior among individuals do not covary with the observable differences in success and prestige? For example, suppose you’re a farmer and everyone in your county uses chemical pesticides, except one farmer who uses natural pesticides and obtains fairly average yields. Do you adopt chemical or natural pesticides? One solution for dealing with such

¹¹ In thinking about content biases, it is important to keep in mind a number of things. First, jury-rigged evolutionary products, like human minds, are likely to contain accidental by-products and latent structures that create biases for fitness-neutral behaviors, ideas, beliefs and values. Boyer (2001) details one kind of by-product content bias in his explanation for the universality of religious concepts (like ghosts). Second, even content biases that arose because they led to the adoption of fitness-enhancing behavior in ancient environments may now promote the adoption of quite maladaptive practices. Third, content biases may be either reliably developing products of our species-shared genetic heritage or they may be culture specific. People may learn valuable content cues via cultural learning. Then, having acquired this idea or practice via cultural transmission, they may be more likely to acquire another, because the two “fit together” in some cognitive or psychological sense.

information-poor dilemmas is conformist transmission: copy the cultural traits of the majority (Boyd & Richerson, 1985; Henrich & Boyd, 1998; Kameda & Nakanishi, 2002). Conformist transmission allows individuals to aggregate information over the behavior of many individuals. Because a person's traits implicitly contain the effects of his individual experiences and social learning (including *his* prestige-biased transmission), conformist transmission can be the best route to adaptation in information-poor environments. To see this, suppose every individual receives an unreliable (but not useless) piece of information from the environment about the highest yielding practice for the current circumstances. This information, for any one individual, might give that person a 60% chance of noticing that chemicals yield slightly larger returns than natural pesticides. Thus, using individual learning alone, individuals will adopt the more efficient farming practice with a 60% chance. But, if an individual samples the behavior of 10 other individuals, and simply adopts the majority behavior, his chances of adopting the superior chemical pesticide technology increase to, say, 75%. By aggregating the partial information of other individuals, conformist transmissions can improve an individual's chances of making adaptive decisions.

Consistent with this theoretical work, a substantial amount of empirical research shows that people do use conformist transmission in a wide range of circumstances, particularly when problems are complex or difficult to figure out on one's own. This work reveals that humans have two different forms of conformity that operate in different contexts. The first, often called *informational conformity*, matches theoretical expectations from models of conformist transmission, and is used to figure out difficult problems and results in people actually altering their private opinions and beliefs about something. The second, often called *normative conformity*, is conformity for the purposes of going along with the group, and not appearing deviant. Under this type of conformity, people alter their superficial behavior, but often don't change their underlying opinions or beliefs. We argue that the ultimate origins of this second type of conformity can be explained by the evolutionary process that we describe under the rubric of social norms, punishment and prosociality, below.

It is now fairly well established that cultural learning is one of our primary means of behavioral adaptation (Alvard, 2003; Tomasello, 1999). Our capacities for cultural learning appear to be adaptations (products of natural selection acting on genes) for acquiring useful behaviors, practices and strategies in complex, information-poor environments. These adaptations can be understood as a set of specialized psychological mechanisms (e.g., prestige and conformist biases) designed to extract a vast array of useful information from other individuals. While themselves part of our genetic evolutionary heritage, they give rise to a second system of inheritance (culture) that evolves in parallel, and in interaction with, our genome. However, culture and cultural learning is only

part of the story. As will be clear in the following section, we also believe that human psychology contains a number of important non-cultural psychological adaptations that deeply influence human social behavior and decision-making, while interacting with ongoing cultural evolutionary processes.

6. Evolutionary theories of cooperation and social psychology

Here we discuss five evolutionary theories that provide potential ultimate solutions to the dilemma of cooperation. Our objectives are to provide the reader with an intuitive understanding of how evolutionary theories can lead to cooperation in humans, and to derive a set of proximate psychological mechanisms and observable behaviors from each theory. Our five classes of evolutionary models are (1) kinship, (2) reciprocity, (3) reputation, (4) social norms and punishment, and (5) ethnicity. In laying these out, we will discuss both some of the fine nuances of how they work (or fail to work) and the role played by culture in both the phylogenetic and ontogenetic (developmental) emergence of these forms of cooperation. Unlike other introductory works on cooperation, one of our main take-home theoretical messages is that all pathways to cooperation require – in humans – some understanding of cultural transmission, either in their operation, or in both their evolutionary origins and their operation. For example, all humans societies have both kinship (psychological and behavioral biases to help genetic relatives), and a culturally-transmitted kinship system that has important effects on individual behavior, but cannot be fully explained by kin selection. Because much of this theoretical material is based on mathematical models that are beyond the scope of this introductory paper, we hope to inspire our readers to learn mathematics and game theory, and explore the primary literature.¹²

6.1. The core dilemma in the evolution of cooperation

There is a simple core principle that underlies nearly all solutions to the puzzle of cooperation: cooperation can evolve under circumstances in which natural selection can take advantage of a stable regularity that allows cooperators to preferentially bestow their benefits on other cooperators. In other words, cooperation can evolve when cooperators tend to cooperate with other cooperators. Eq. (1) expresses this condition in its most general sense (see Henrich, 2004):

$$\beta b > c. \quad (1)$$

Here c is the cost paid by the cooperator in order to deliver an amount of benefits, b , to another individual or group of individuals. These costs and benefits are measured in units

¹² Henrich (2004) and associated commentaries provides an entrée into the primary theoretical literature.

of fitness (e.g., number of offspring). β is a statistical relationship called a regression coefficient. It measures the degree to which ‘being a cooperator’ predicts ‘bestowing benefits on other cooperators’. In the simplest case, it is the probability that a cooperator is bestowing benefits on another cooperator. When this inequality is satisfied, natural selection can favor the spread of genes that build proximate psychologies for cooperative behavior. All solutions (and their psychological incarnations) must be able to create, or at least sustain, the statistical regularity that allows cooperators to benefit other cooperators. Interestingly, in the history of the study of the evolution of cooperation (Frank, 1995, 1998), this simple equation was first derived for specific cases (e.g., kinship and reciprocity) before this more general, abstract condition was derived.

6.1.1. *The rise and fall of green-bearded cooperators*¹³

Once this general problem is understood, a simple solution suggests itself. Imagine a gene that causes its bearer to both ‘have a green beard’ and to only help other green-bearded individuals. In our equation above, β would be at its maximum value of one (greenbeards only deliver benefits to other greenbeards) and cooperation would spread rapidly. Soon the entire world would be green-bearded cooperators, and everyone would be merrily cooperating with everyone else. However, the statistical relationship (represented by β) between bestowing help and being a cooperator must be reliable and durable. Now imagine we are in a jolly world of green-bearded cooperators and a mutant green-beard emerges. This fellow has the requisite green beard, but he’s not a cooperator. He never helps anyone, but everyone always help him because of his lush, green beard. Consequently, the mutant, and his mutant gene, will be very successful, and will eventually drive the green-bearded cooperators to extinction. The world now consists entirely of green-bearded defectors. The trick to solving the dilemma of cooperation is not so much in producing a positive β value (i.e., in creating a way for cooperators to find other cooperators), but in figuring out how to maintain a reliable, stable, β value.¹⁴ Kinship provides one set of avenues to cooperation.

6.2. *Kinship*¹⁵

Let’s begin with one of the most prevalent forms of cooperation in nature. Consider a mother with a gene (or genes) that cause her to experience deep positive emotions toward her offspring (e.g., baby, egg), and these emotions

cause the mother to give help to her offspring. Why would this gene spread? Following the general logic outlined above, it could only spread if the recipient of the ‘help’ was also a ‘helper’. That is, if the offspring is also a carrier of the gene the leads to a psychology that evokes the rule ‘help your offspring’. This leads to the question: What are the chances that her offspring has the helping gene given that the mother has the helping gene? The answer can be rather complicated, but if the organism has a genetic system like humans (diploid), then the chance should be at least 50%. The main reason is simple. Mother’s offspring will on-average share 50% of her genes, which means that there is a 50% chance that the offspring has a copy of the mother’s ‘help your offspring’ gene. Thus, the characteristic of ‘being one’s offspring’ is a reliable predictor (has $\beta = 0.5$) of sharing the relevant cooperation gene(s). This kind of evolutionary process is called ‘*Kin Selection*’ because mother and offspring (or any blood relatives) share copies of the *same gene by descent from a recent common ancestor*. In a sense, by causing the mother to help her offspring, the gene is helping itself to survive and reproduce down the lineage.

To illustrate the crux of the dilemma of cooperation, suppose a variant of the ‘helper’ gene just described mutated into existence and caused its bearer to feel equally affectionate towards all juveniles, not just her own offspring, and thus to direct help toward whomever most needed it. Could this mutant gene spread? No, because offspring of the original gene, ‘help your own offspring’, would not only get help from its own mother, but also help from other mothers carrying the mutant gene, while offspring of mothers with the mutant gene would get only minimal help from their own mom (who would be spreading their help around), and no help from mom’s carrying the ‘help your own offspring’ gene. This would, on-average, lead the ‘help your own offspring’ gene to outcompete the ‘help everyone’s offspring’ gene.

Returning to the Core Principle, natural selection will favor the evolution of psychological mechanisms that allow cooperators to focus their benefits on individuals who are likely to be cooperators. From this perspective, kinship represents a class of different proximate mechanisms that take advantage of the fact that some individuals in any population will tend to share the same genes *by descent from a recent common ancestor* (like a parent). This fact of biology creates numerous evolutionary opportunities for natural selection to find reliable statistical regularities to exploit. ‘Blood relatives’ share many other characteristics, besides the all-important cooperation gene(s), that natural selection can take advantage of in building psychologies that help in directing benefits at other individuals likely to share cooperation genes. For example, close kin, such as siblings, may share a similar appearance or smell that natural selection may use as cues to build a psychology like ‘help those who look and smell like you’. These similarities may be related to sharing some of the same genes (ones not related to cooperation), or they may be related to

¹³ Dawkins (1976) coined the colorful “greenbeard” example.

¹⁴ Interestingly, ‘greenbeard solutions’ constantly re-emerge in the literature on the evolution of cooperation. Unknowingly, researchers continually re-produce ‘solutions’ to the evolution of cooperation that are actually green-beards in disguise. Having now slandered the green-beard solution, it must be noted that there is at least some evidence for greenbeards in nature (Keller & Ross, 1998).

¹⁵ Key references include Hamilton (1964, 1972), Grafen (1985), Frank (1997) and Queller (1992).

having been reared by the same female in the same environment (e.g., nest). Even proximity can provide an evolutionary opportunity to construct a psychological mechanism capable of maintaining stable cooperation (e.g., ‘help those who tend to hang around mom’). It is important to realize that natural selection does not ‘care’ about how a particular cue arises (i.e., it does *not* care about kinship per se), only that the cue is a stable and reliable predictor of who is likely to be a cooperator. It so happens that in a variety of species, many aspects of reproduction create reliable statistical patterns on which natural selection can build. However, the details always depend on the specifics of the particular species. Natural selection might, for example, lead to a system that causes newborns to smell those around them during the first few days after birth, and subsequently direct benefit towards those individuals for life. This species must have a social structure that reliably places newborns among close kin, a physiology such that either individuals or close kin produce distinct (distinguishable) scents, and some way to exclude non-kin from sneaking into the nest during those first few days or faking the scent of other family members – the ‘exclusion’ need not be fool proof for some cooperation to evolve, but the better it is, the more cooperation can evolve.

In studying how kinship can solve the dilemma of cooperation, Hamilton (1964) derived the simple rule that now bears his name (Hamilton’s rule):

$$rb > c. \quad (2)$$

The reader should note the similarity between Eqs. (1) and (2). As in (1), b is the benefits bestowed and c is the cost to the bestower. Now, however, r replaces β and represents the ‘coefficient of relatedness’, which specifies both the average proportion of genes shared by the two individuals as a consequence of recent common descent, and more importantly, the probability that the receiver of help shares a specific gene (i.e., cooperative gene). The parameter r is a special case of β that occurs when the foundation of cooperation is built on kinship and an organism can, by some means, direct benefits preferentially toward their kin.¹⁶

If we assume that by using a variety of these cues, humans (and other animals) can assess their degree of relatedness, then we should be able to predict who they will be most likely to help. This suggests that individuals should

cooperate with relatives according to their ‘coefficient of relatedness’ (r), which is the probability that they share the same gene by descent. In our species, siblings and parent-offspring have the highest r at 0.50. Assuming only one line of descent (no interbreeding) grandparents and their grandchildren and half-siblings have the next highest at $r = 0.25$. First cousins are related at $r = 0.125$. Using the evolutionary logic of kinship, parents and their children, and siblings, should cooperate a fair amount, and first cousins may cooperate a tiny bit, but more distant relatives have too little relatedness to cooperate with each other. Among siblings, for example, the fitness cost (in terms of survival and reproduction) must be less than half of the fitness benefits delivered before cooperation would be favored. To get an intuitive sense of this, consider to whom you would give a kidney. In the US, 68% of kidney donations come from kin, while less than one-half of 1% come from anonymous strangers (and thousands die every year waiting for kidneys).

This body of theoretical work predicts that humans will likely possess a “kin psychology” that is designed to (1) identify kin and (2) direct benefits toward close kin. This should be particularly true when the costs of cooperation are relatively high compared to the expected benefits to the person helped. The available evidence suggests that humans may use a variety of cues to assess kinship, including physical similarity to themselves and other family members, scent, proximity during youth (Wolf, 1995), food sharing, and social learning (by watching and listening to others) to ‘figure out’ to whom they are related, and thus who is more likely to be a ‘cooperator’. These cues affect affective states or feelings that, in turn, influence (along with many other factors) cooperative behavior and trust with cued individuals. For example, in experiments measuring trust using allocations of real money, researchers have shown that people are more trusting of individuals who resemble themselves. The researchers varied ‘resemblance’ by using an image-morphing technology to combine an image of the subject with another person (DeBruine, 2002). This suggests that physical resemblance to self may cue affective states built by natural selection to benefit kin. Using the same kind of technology, other researchers have shown that men are more kindly disposed towards babies that resemble themselves (Platek, Burch, Panyavin, Wasserman, & Gallup, 2002; Platek et al., 2003). This is important because men, unlike women, can rarely be completely positive that they are a child’s genetic father. While not focused on cooperation, a variety of other studies show that our evolved psychology uses scent and early-life proximity to calibrate affect and target behavior.

In our book on cooperation (in press) we show that Chaldean immigrants in Detroit restrict their most costly forms of cooperation to close relatives, thereby confirming a standard finding from small-scale societies (Chagnon & Irons, 1979; Cronk, Chagnon, & Irons, 2000). However, this work yields two additional insights. First, while Chal-

¹⁶ It is a common misconception that kinship depends on sharing a certain percent of the same genes, and that r gives the fraction of genes shared by two relatives. This is false for two reasons. First, natural selection will favor genes that direct benefits at identical copies of themselves, not other genes, so percent of shared genes is theoretically irrelevant. Thus, r should be thought of as the probability that another individual has a copy of the ‘helping gene’ given that the first individual has it. Second, humans already share most of the same genes because we are the same species. Sharing genes for building finger and blood vessels is not important for understanding human cooperation. Thus, r does not give the percent of shared genes; it does happen to correspond to the percent of genes that are identical by descent from a recent common ancestor. But, as we said, this is only relevant in that it may create reliable patterns that natural selection can exploit.

deans have an explicit cultural ideology that all Chaldeans are “one big family” and that everyone is related, no one is fooled behaviorally and the lines of costly cooperation are clearly drawn at the outskirts of the immediate family. This finding bears on the claim made by some evolutionary psychologists that large-scale human cooperation in contemporary societies results from the faulty (maladaptive) operation of a psychology designed for cooperating in the kin-based societies that many believe characterized human ancestral societies.¹⁷ At least among the Chaldeans of metro-Detroit, there is no confusion about kin and no evidence for misdirected cooperation, despite the fact that they live in a large urban center. Thus, while kin psychology can explain cooperation among closely related kin, it does not explain cooperation in other contexts.

The Chaldeans also illuminate the issue of kinship and cooperation in a second way. Our findings show kinship and the cultural expectations about behavior (carried from Iraq) combine to allow many Chaldean storeowners to out compete non-Chaldeans economically. Moreover, it appears that Chaldean cultural beliefs are currently shifting toward models typical of other middle-class Americans. The case material suggests that while kin psychology is constant, adaptive cultural learning processes are gradually shifting cultural beliefs, which is altering the details of ‘who helps whom’. This shift has important economic, social and educational impacts for Chaldeans. That is, culture matters.

6.3. Reciprocity

Reciprocity represents another well-studied class of potential solutions to the evolutionary puzzle of cooperation (Trivers, 1971). Reciprocity works on the commonsense logic of “if you scratch my back or someone else’s, I’ll scratch yours”. Individuals, by applying such tit-for-tat reciprocal strategies, can preferentially associate themselves with other cooperators, and thereby increase both their chances of bestowing their benefits on other individuals with the same genes, and by increasing their likelihood of receiving benefits from others by putting themselves in the company of other reciprocators.

Progress on reciprocity as the answer to the dilemma of cooperation has differed from kinship in two important (and often unrecognized) ways. First, despite literally hundreds of papers on the topic (Axelrod & D’Ambrosio, 1994: shows 209 publications from 1987 to 1993), the theoretical conclusions derived for the mathematical models and computer simulations are substantially more ambiguous, nuanced and qualified than those for kinship. Second,

the empirical evidence for reciprocity-based cooperation in *non-human* species is scant (Hammerstein, 2003), especially when compared to the evidence for kin-based cooperation. Nevertheless, in our species, reciprocity consistently re-emerges from both ethnographic and experimental studies.

Below, we deal with these two aspects of reciprocity by synthesizing the qualitative findings from a substantial body of theoretical work and linking these to our broader empirical efforts. Unlike most other general treatments, we will set this work within the general framework of the Core Dilemma, and integrate it with humanity’s reliance on cultural learning. This last connection allows us to explain why reciprocity-based solutions are rare in nature while being plentiful and diverse in human societies. Theoretically, there are two take-home messages here: (1) reciprocity-based solutions to cooperation are much less robust than many scholars think, and (2) these ‘solutions’ are so prominent in humans *because of* – not *despite* – our evolved cultural learning capacities. Empirically, we use the qualitative insights developed below to generate a series of predictions about how our evolved ‘reciprocity psychology’ works. In our book on cooperation, we bring a range of experimental and ethnographic data to bear on many of these predictions.

Our general theoretical framework shows that the evolution of cooperation requires that benefits be preferentially bestowed on cooperators. Above, we explained how kinship can facilitate this by providing reliable ‘cues’ for identifying cooperators. Reciprocity-based solutions, however, use different mechanisms to support cooperation among non-kin. In this class of solutions, natural selection favors individuals who can use the past behavior of other individuals as an indicator of whether they are a cooperator or not. If a person’s past behavior suggests that he may be a cooperator (regardless of what genes he carries!), then natural selection should favor a psychology that promotes bestowing benefits on that person. There are two reasons for this selection pressure. First, an individual’s past history may act as a costly cue that he/she carries a ‘cooperative gene’. By bestowing benefits on those with a history of cooperation, cooperators may be able to preferentially direct benefits toward others with the same gene. Second, bestowing benefits on certain individuals can cause a return flow of benefits back to the bestower. In this case the specific genes carried by the other individual are irrelevant. What matters is that bestowing benefits causes a return flow of benefits back to the one with the ‘cooperation gene’. There are two sources of information about past behavior that might diagnose on whom one should bestow benefits. These sources lead to the two commonly discussed forms of reciprocity: (1) *Direct reciprocity*, based on direct, personal experience with the other individual(s), and (2) *Indirect reciprocity*, which involves getting information about potential interactants by observing them with others, or by gathering reputational information (culturally transmitted information) about their past behavior with other individuals.

¹⁷ This hypothesis for explaining large-scale human cooperation is the kinship version of the “Big Mistake Hypothesis” (Boyd & Richerson, 2002b). Despite being repeatedly criticized on both empirical (ethnographic and experimental) and theoretical grounds (Boyd & Richerson, 2002b; Fehr & Henrich, 2003), the Big Mistake Hypothesis remains widely believed in many corners of evolutionary scholarship.

6.3.1. Direct reciprocity

The intuition behind direct reciprocity is simple: If you help me, I will help you. If you stop helping me, I will stop helping you. Direct reciprocity (DR) depends on direct, ongoing experience between interacting individuals. To illustrate, suppose Joe and Natalie are each small business owners. Each business goes through an annual period of a cash shortfall at different times of the year. If each business can get through the period of cash shortfall then each can make an overall profit, but if it cannot, the business makes no profit. Getting through the period of scarcity requires \$100. If both Joe and Natalie give each other money when the other needs it, both make \$300. If neither gives money to the other, both make zero profit. If Natalie gives Joe money during his scarce period, but Joe decides not to give Natalie any money during her scarce period, Joe will make \$400 (\$300 plus the \$100 he did not give out = \$400) and Natalie will lose \$100 (zero profit minus \$100 given to Joe) (see Payoff Matrix, below). By ‘defecting’ (not giving the money), Joe does relatively better than Natalie (by \$500) in that year, and Natalie may go out of business, leaving Joe’s ‘business practices’ to proliferate as people see his financial success and copy his strategy. Over time as more and more people act like Joe, no one will give any money and everyone will go out of business. If Joe and Natalie had sustained cooperation then the pair would have done much better overall than they did when Joe defected. Had they sustained cooperation, after 10 years each would have \$3000 profit. Instead, Joe made \$500 the year he defected but never made anymore in subsequent years because his trading partner, Natalie, went out of business and lost \$100. Over 10 years, their combined profit was only \$400.

Payoff Matrix

Prisoner’s dilemma	Cooperate (give money)	Defect (don’t give money)
Cooperate	300, 300	–100, 400
Defect	400, –100	0, 0

To study the kinds of strategies or behavioral rules that will succeed in maintaining cooperation through reciprocity, theorists have formalized the above situations in an abstract format called the Iterated Prisoner’s Dilemma (IPD). This mathematical formalization allows researchers to systematically study the properties of various strategies, and other variables, in solving the dilemma of cooperation. In the IPD, individuals are paired at random, interact repeatedly for some expected number of rounds (like Joe and Natalie giving money to each other each year), and receive payoffs based on their decisions and those of their partner. Individuals using strategies that give them higher payoffs relative to other strategies will increase in relative frequency compared to those using strategies that yield lower payoffs. The game matrix illustrates the payoffs, and captures Joe and Natalie’s business situation.

Some of the earliest work by Axelrod and Hamilton on direct reciprocity using this approach suggested that simple reciprocating strategies could generate long-term, stable, cooperation (Axelrod & Hamilton, 1981). One of the central analytical findings in Axelrod’s book (1984), *The Evolution of Cooperation*, shows us how repeated interaction with the same player is merely another way to address the Core Dilemma in the evolution of cooperation (also see Maynard Smith, 1982). If we take a simple reciprocating strategy that cooperates in the first round and then in subsequent rounds does whatever the other player did in the previous round (TFT = tit-for-tat) and assume that this strategy (TFT) is common in a population, the condition for the reciprocating strategy to remain common against low frequency invading defectors is

$$\omega b > c. \tag{3}$$

Here, as above, b and c are the costs and benefits of cooperation, but now ω (replacing β from (1)) is the probability that the interaction with the same individual will continue to the next round. Essentially, ω is a measure of how long the cooperation can continue if both individuals keep cooperating. In this way, prior interactions with the same person in a repeated sequence of interactions provide a ‘cue’ about who to preferentially interact with, so as to receive a flow of benefits. The longer the interaction continues with the same reciprocator, the greater the amount of cooperation that can be sustained, because the individuals receive more total benefits.

While several of the general findings put forth by Axelrod have held up well (we discuss those shortly), many of these findings have not stood the test of time, and have been over interpreted by others. Perhaps the starkest example of this is the reciprocating strategy “tit-for-tat” (TFT), which was thought to be the most robust and simplest solution to the problem of cooperation. All the hoopla surrounding TFT resulted in a number of misunderstandings, and led a well-know evolutionary biologist to write in a textbook that TFT “is superior to all other [strategies] in playing repeated games of prisoner’s dilemma” (Trivers, 1985, 391). Not only was such a statement not supported by the existing work at the time, but soon after 1985 a long parade of papers showed that TFT wasn’t even ‘pretty good’ in many situations.¹⁸ For example, in environments in which the information about the payoffs or behavior of one’s partner is less than perfect, or simply in larger groups, TFT does not lead to very much cooperation and there are much better strategies. In these and other situations, there is a multiplicity of other ways to preserve cooperation, but this diversity in itself creates an evolutionary challenge that only culture can solve.

¹⁸ Readers interested in exploring all the ways in which TFT fails should begin with papers by Bendor, Kramer, and Stout (1991), Bendor, Kramer, and Swistak (1996); Boyd (1989, 1992); Boyd and Lorderbaum (1987); Hirschleifer and Martinez-Coll (1988).

Overall, the important intuitive understandings about reciprocity that have arisen since the mid-1980s suggest that (1) reciprocity-based solutions are substantially more complicated than many thought and (2) while these complicated, contingent patterns lack the desired simple elegance sought by many, these appear more consistent with empirical patterns among human societies and across species. Below, we synthesize what is known about reciprocity in an effort to paint a coherent image about the evolution of cooperation via direct reciprocity. Using this image, we sketch the entailments for our evolved psychology and human behavior.

On the positive side, while no simple strategy exists that yield anything approaching a robust *reciprocity-based* solution to the dilemma of cooperation, much of this theoretical work suggests that some types of reciprocating strategies can lead to substantial amounts of cooperation in a wide range of circumstances, as long as group size remains small (cooperation occurs in dyads, triads, etc.) and interaction endures for a sufficiently long time (large ω). Yet, while some kinds of reciprocating strategies can be successful in particular circumstances, the details of such strategies vary tremendously. To give a sense of which aspects of the cooperative environment influence the success of different kinds of reciprocating strategies, we review four key factors: *group size*, *noise*, the *ecology* of other strategies, and variation in the *length of interaction* across partners.

6.3.2. *Direct reciprocity: does group size matter?*

The success of TFT and other reciprocity-based strategies in pairwise interactions (interaction between only 2 individuals at a time) in much of the early work led many to assume that these 2-person findings would extend to explain cooperation in larger groups – what are called *n*-person cooperative dilemmas or public goods problems. Rather than two individuals helping one another, the *n* players in public goods games contribute to the benefit of everyone in the group. If one assumes that individuals occasionally make mistakes (defecting when they meant to cooperate), then reciprocity-based strategies (like TFT) do not generate cooperation in larger groups to nearly the degree that they do in 2-person cooperative situations.¹⁹ The capacity for reciprocity to maintain cooperation decreases geometrically with increasing group size. At the sociological level, this line of theoretical work suggests that we should not expect direct reciprocity to be the primary factor in maintaining cooperation in large groups. At the psychological level, this finding predicts that human psychology should be motivated by direct reciprocity *only* when the cooperative unit is small – dyads should be the preferred group for direct reciprocity. From the per-

spective of institutional design, our direct reciprocity psychology will be most effectively employed when the cooperative group is small and enduring.

6.3.3. *Direct reciprocity: what if information about payoffs or the other people's behavior is noisy?*

Reciprocity-based strategies of all kinds are entirely reliant on the information that people receive about the past behavior of their partners. Given this, what if individuals receive inaccurate information about what their partners did, or what payoffs they received in previous interactions? This is certainly a potential problem in the real world, as information is often ambiguous or uncertain.²⁰ Continuing with the example from above, what if Natalie figured that Joe had decided not to give her the money, but actually Joe's check got lost in the mail? Joe tried to cooperate but Natalie perceived him as a defector. Or imagine a case in which your partner helps you move furniture for the full 6 hours that you wanted him, but due to some confusion, you thought he was only there for 4 hours (and you helped him 6 hours last time)? What do you do when he asks you for 6 hours of help next time? Do you stop reciprocating entirely (perhaps you think he's taking advantage of you), or do you reciprocate only 4 hours (playing TFT)? Or, do you help him for the full 6 hours?

Noise in this kind of information creates a tradeoff between a strategy's PROVOKABILITY and a strategy's GENEROSITY. PROVOKABLE strategies show a hair-trigger willingness to stop cooperating as soon as a person has any indication that her partner is not cooperating. GENEROUS strategies reciprocate with more than they are given – so if you give me 4 hours of your time (or at least I think that you do), and I give you 6 next time, then I'm GENEROUS. TFT, for example, is quite PROVOKABLE but not GENEROUS at all. This is because a person using TFT immediately stops cooperating when her partner appears to have defected. Being PROVOKABLE is important to avoid being exploited by defecting strategies that take advantage of noisy environments. TFT is not GENEROUS because TFT gives back whatever its partner appears to have given. Not being GENEROUS in a noisy environment can lead to substantially reduced amounts of cooperation, as misunderstandings lead to irredeemable losses in cooperation.

To see this, suppose two TFT were interacting – let's call them TFT1 and TFT2. On round one TFT1 gave 10 (the maximum) to TFT2, but due to the noisy environment, TFT2 thought he was given only 8. Then, in the next round TFT2 gave 8 to TFT1, but TFT1 thought he had received 7, and he returned 7. Now, while some misunderstanding can drive exchanges back up, the average exchange would not be near 10 – actually, this setup leads to an average exchange of 5 in the long run. In contrast, a GENEROUS strategy can deal with this. Suppose two TFT + 2 strategies

¹⁹ The relevant theoretical findings were independently arrived at by different researchers around the same time, using somewhat different models (Bendor & Mookherjee, 1987; Boyd & Richerson, 1988; Joshi, 1987).

²⁰ All of Axelrod's early work was in a "noiseless environment" in which actors always had perfect information about their partners' past behavior.

(TFT + 2 gives back two more than it was given) are playing in a noisy environment. If TFT + 2[1] gives 10 to TFT + 2[2], but TFT + 2[2] believes that only 8 was sent, he will still give 10 back to TFT + 2[1]. Careful analysis has shown that this generous approach leads to more cooperation in a noisy environment than a PROVOKABLE strategy. However, the effectiveness of being GENEROUS depends critically on what other strategies are lurking out there because generous strategies that can maintain effective cooperation in noisy environments are susceptible to exploitation by crafty strategies that hide under the cover of noise and exploit the GENEROUS strategies. If you are in a nice environment with mostly reciprocating strategies, GENEROUS makes the most of cooperation. But, if exploiter strategies emerge in sufficient numbers, being PROVOKABLE is the only defense. Interestingly, GENEROUS strategies can thrive if there are a sufficient number of PROVOKABLE strategies to keep the crafty defecting strategies at bay. GENEROUS strategies like TFT + 2 bring out the best in PROVOKABLE strategies like TFT. In noisy environments, TFT cooperates at much higher levels with TFT + 2 than with another TFT and TFT keeps lurking exploiter strategies at bay.²¹ *Strategic diversification can help immunize a population against invading defectors and thereby promote cooperation by reducing the inherent tradeoffs between being PROVOKABLE vs. GENEROUS.*

Besides favoring generosity, noisy environments can also favor CONTRITION. Contribute or remorseful strategies accept punishment (in the form of a defection), after they have defected. CTFT (Contribute TFT) cooperates in the first round and then reciprocates cooperation and defection in subsequent rounds. But, after a mistaken defection by CTFT, if CTFT's partner defects then CTFT will cooperate in the next round despite its partner's defection. After this, CTFT will return to playing TFT. This allows CTFT to repair cooperative relationships that would otherwise be destroyed by some kinds of errors (Boyd, 1989). For example, if CTFT is interacting with TFT and CTFT accidentally defects when he meant to cooperate, he will 'accept' TFT's defection on the next interaction and cooperate one time after this. This cooperation will bring out cooperation from TFT and the pair will return to cooperative interaction. If two TFTs were interacting, all cooperation would have ceased.

However, contrite strategies are susceptible to 'error in perception' (Boerlijst, Nowak, & Sigmund, 1997). With these errors, there arises a mismatch between what the individuals in the interaction believe happened. For example,

Natalie might believe Joe defected when, in fact, Joe cooperated and believes he cooperated. Even if Joe is using a contrite strategy, the pair won't be able to repair their relationship because Natalie will defect in the next round, and Joe won't understand why (because he thinks he cooperated). Contrition cannot always save you from errors.

One final aspect of noisy environments is the potential importance of remembering more than just the previous interaction – having a *longer-term memory* for previous interactions. While some kinds of memory can be an effective means to sustain cooperation in noisy environments, longer memories are not necessarily better and can actually lead to less cooperation. Longer memory can lead individuals to adapt themselves to the *noise*, rather than the opportunities for cooperation available from the array of other strategies in the ecology. Additionally, more memory is cognitively costly, especially as a person's number of partners increases (Bendor, 1987; Bendor, 1993). The importance of memory in sustaining cooperation via reciprocity is often greatly overrated. It has been frequently assumed that humans cooperate more than other animals because they can remember a longer history of interactions. This is a case in which common intuitions fail in the face of rigorous formal models.

6.3.4. Direct reciprocity: what if the length of interaction varies across different individuals?

Under a wide range of conditions, reciprocating strategies that lead to long-term cooperation are generally NICE. That is, successful strategies generally cooperate in the first round of an interaction. Non-NICE reciprocators, like suspicious tit-for-tat (STFT) who defects in the first round and then plays TFT, can often spread and drive out NICE reciprocators. But, usually, STFT does not create much long-term cooperation when it's common. Consequently, being NICE should depend on cues about how long interactions might go on, and whether there are other reciprocating strategies out there to cooperate with – if everyone is a defector, being NICE does nothing for cooperation. One implication is that in a population that has individuals who vary in how long they stay around (different ω 's), and this difference cannot be easily detected, then non-NICE strategies who wait until they are sure they are with a long-term interactant before initiating cooperation can be the most successful. By waiting, non-NICE strategies can effectively target their cooperation at other long-term interactants. This suggests that natural selection should favor individuals who are (1) NICE to long-term interactants, (2) non-NICE and non-cooperative to short-term interactants when such individuals can be distinguished from long-termers, and (3) initially wary (non-NICE) but eventually cooperative when the population is a mix of long-term and short-term interactants and the types cannot easily be distinguished (Boyd, 1992).

It has been argued that humans, as a consequence of our long history in small-scale societies, would not have experienced enough ephemeral interactions that had important

²¹ These insights were gleaned from a variety of sources (Bendor, 1993; Bendor et al., 1991; Bendor & Swistak, 1997; Boyd, 1989; Hirshleifer & Martinez-Coll, 1988; Wu & Axelrod, 1995). Another interesting aspect of generosity in noisy environments is that the most successful strategies overall may actually lose to every other strategy in direct pairwise competition. Similarly, a strategy that defeats every other strategy in pairwise competition will often place dead last when a variety of strategies are let loose together.

fitness consequences for our psychology to be calibrated to differentiate short-term interactants from long-term interactants. This argument leads to the claim that people (unconsciously) assume that everyone is a long-term interactant. That is, even though we can consciously recognize that we aren't likely to meet someone again, we cannot really think that way. This version of the "Big Mistake Hypothesis" has numerous problems, but first among these problems is that the ethnographic, ethnohistorical and archaeological evidence indicates that people in small-scale societies routinely have important (i.e., fitness relevant) interactions that are short-term or one-shot. The Big Mistake Hypothesis is rooted in a false, but widely believed, anthropological myth about the nature of life in small-scale societies. There is, in fact, every reason to believe that ancestral human life involved fitness-relevant interaction with short-term interactants (Fehr & Henrich, 2003).

6.3.5. *Direct reciprocity: changing ecologies of strategies and the real challenge of reciprocity*

Above we have hinted that the success of a strategy depends on the other strategies that are in the mix (exist in the population). This turns out to be a remarkably general property of reciprocity-based cooperation. For any strategy one can devise, there is a combination of other strategies that will destroy it! In a variety of situations it has even been possible to prove (mathematically) that no reciprocating strategy is safe and robust against other combinations of strategies (Bendor, 1993; Bendor & Swistak, 1997; Boyd & Lorderbaum, 1987; Farrell & Ware, 1989; Lorberbaum, 1994; Lorberbaum, Bohning, Shastri, & Sine, 2002). This means that successful individuals need to be able to rapidly adapt their strategies to the changing balance of other strategies in the population. This presents a particularly prickly problem for genetic evolution and natural selection to solve, since shifts in the social–ecological balance may occur rapidly, even within the lifetime of an individual. Because there are so many possible social ecologies, individuals in many species (especially humans) are likely to encounter numerous social ecologies that their species has never encountered in the evolutionary past. As we'll explain, we believe that the human adaptation for cultural learning provides one of the few means through which natural selection can meet this – the real challenge – of the evolution of cooperation via direct reciprocity.

6.3.6. *Direct reciprocity: why is reciprocity-based cooperation rare in non-humans?*

The impact of small shifts in the ecology of other strategies in the environment, as well as the importance of variables like uncertainty and group size, have been underappreciated by many evolutionary scholars. We think that the right take home message from all this theoretical work is that the genetic evolution of cooperation via direct reciprocity is **not** a robust solution for most animals in most complex exchange contexts. In a sense, the theoretical work shows that there is no way to build an 'all-purpose

reciprocity machine'. Even if one could somehow encode the rules for several different reciprocity strategies into the genes of an organism, there is a virtually infinite number of strategies one would need. This creates a combinatorial explosion of potential contingencies that would have to be built into the genes.²² The situation is especially acute in humans, given the variety of cultural and physical environments in which humans operate.

Consistent with the above logic, the empirical record shows little evidence for reciprocity in non-humans (Hammerstein, 2003), and the non-human examples that do exist are relegated to special case situations.²³ In nature, and in general, reciprocity is simply not a wide-spread and robust solution to the dilemma of cooperation in the way that the kinship solution is robust.

6.3.7. *Direct reciprocity: enter cultural transmission*

However, while this line of theoretical argument is consistent with non-human data, we are left with a puzzle in the human case. In contrast to non-humans, the available ethnographic and empirical evidence shows that direct reciprocity is a strong and recurrent pattern among our species, and across cultures. Yet, interestingly, the details and circumstances of direct reciprocity are highly variable across contexts and human social groups (Fiske, 1991). We argue that the combinatorial explosion is solved in humans through cultural transmission. The adaptive nature of cultural learning works with a psychology for reciprocity (that coevolved with culture) to create 'custom-fit' strategies (solutions) that are dynamic and adaptable to the challenges of reciprocity described above.

To understand how culture solves the combinatorial explosion of reciprocity, it is important to recognize that (1) cultural evolution occurs at much faster rates than genetic evolution (e.g., novel practices can spread through a population in a single generation: Rogers, 1995), (2) such cultural learning processes allow populations to rapidly adapt to novel situations without any genetic change,²⁴ and (3) cultural transmission can, to a degree, construct its own environment (e.g., written records reduce the 'noise' in payoffs between interactants, or institutional structures can turn *n*-person public goods problems into dyadic interactions). Genetic evolution will continually

²² Those familiar with evolutionary psychology might note that this is the same 'combinatorial explosion' logic used to defend the theoretical claim that our cognition is composed of many special-purpose modules. Thus, the same logic that leads one to believe the mind is modular also leads one to be rather suspicious of arguments that humans have 'reciprocity psychology' that operates *without* adaptive cultural input.

²³ We think the intuitive appeal of TFT clouded what should have been an evident puzzle: if TFT is so simple and robust, why don't we see more cooperation of this type in nature? Answer: TFT isn't that robust, so we shouldn't see much reciprocity-based cooperation.

²⁴ Most of human adaptation to physical environments is cultural. Blow guns, kayaks, boomerangs, bone tools, and poison arrows are all adaptations in the classic sense, and almost entirely culturally learned (Henrich, *in press*).

introduce novel mutations (that lead to new strategies), but cultural evolution will be able to counter such mutations by building new culturally-evolved strategies, or altering the social environment by building new forms of organization (e.g., ID cards, police forces, neighborhood watch), at a rate much faster than genetic evolution. Cultural learning processes like prestige-biased transmission will allow individuals, within their lifetime, to adapt their reciprocity-based strategies to shifting social and physical ecologies. The effect of cultural evolution on reciprocity-based cooperation accounts for the empirical fact that humans in different societies use reciprocity to differing degrees and in different contexts. Our approach can account for *both* the presence of local variation and the universal aspects of human reciprocity.

While genetic evolution alone cannot provide a solution to the real challenge of direct reciprocity, it can facilitate the success of an individual by providing learning mechanisms, biases, and default settings that allow individuals to rapidly and effectively acquire locally useful strategies. We suspect that our ‘reciprocity psychology’ is a product of the coevolution between genes and culture, and rather than supplying rules like TFT, our psychology provides the machinery for rapidly learning the locally-successful rules of reciprocity from other people and from experience. Following from our prior discussion of cultural learning, by imitating successful individuals, people can acquire the locally adaptive strategic nuances that fit the local ecology of strategies, noise, etc. Projecting back into evolutionary time, we might imagine that after our capacities for cultural learning evolved genetically then cultural evolution produced the first simple rules of reciprocity (via imitation²⁵), in the same fashion as cultural evolution produced early stone tool traditions. Following on the heels of cultural evolution, genetic evolution via natural selection might have favored ‘learning genes’ that allow naïve individuals to more rapidly acquire reciprocity-based ‘rules’ (strategies). With these new learning biases in place, cultural evolution may have been able to solve an increasing number of cooperative problems using a variety of different reciprocity-based strategies in different contexts and places. The outcome of this interactional process would be an evolved reciprocity-psychology that operates through (at least partially), and is dependent on, cultural learning to acquire the adaptive strategies for social behavior. This evolved psychology should assist individuals in acquiring effective strategies, avoiding exploitation, and identifying good long-term partners. In a sense, genetic and cultural evolution have teamed-up to solve a variety of cooperative dilemmas through the logic of direct reciprocity.

6.3.8. *Patterns of direct reciprocity*

Based on intuitions that we have built up from reviewing these many theoretical findings on reciprocity, we now have an idea of the general shape of this cognitive and emotional adaptation. Because our evolved ‘reciprocity psychology’ coevolved with culture, most of what’s below can be understood as learning biases that assist individuals in acquiring successful strategies in various ways. Many aspects of these psychologies can be thought of as learning biases that favor the acquisition of certain strategies in some circumstances or as inferential rules that assist individuals in figuring out what underlying strategies a successful person might be applying – remember, people only see the overt behavior of others, and they must infer the underlying strategy in order to acquire it. Such biases, in combination with other cultural learning mechanisms like prestige-biased transmission and conformist transmission, allow individuals to adapt and calibrate their strategy to the local situation. While these biases exist in the minds of individuals, they have a direct impact on the kinds of reciprocity we should expect to observe in groups. Here, we enumerate some of the psychological predictions derived from the available theoretical work.

- (1) **NICENESS.** How nice should an individual be in the beginning of a potentially longer-term cooperative interaction? When groups are large ($n > 5$), the expected length of the interaction is short or highly variable, or the ratio of benefits to costs (b/c) is small, individuals should tend to adopt **SUSPICIOUS** strategies (strategies that aren’t **NICE** in the first round, but can be persuaded to cooperate by some **NICE** strategies). Elsewhere (Henrich & Henrich, *in press*), we have discussed how cultural evolution can facilitate **NICENESS**, and thus promote sustained cooperation, by (1) turning an n -person cooperation dilemma into a dyadic situation, (2) guaranteeing that individuals will stick around for the long run, and (3) using kinship to transcend the life of an individual and extend the time horizon into the future (using kinship to make ω large: families can cooperate over generations). We will also explain how culture can solve other aspects of **NICENESS** by setting standards for ‘how nice to be on a first interaction’ – without culture this is completely unspecified, and there are no standards or expectations.
- (2) **GENEROUS/CONTRITE VS. PROVOKABLE.** How **GENEROUS**, **CONTRITE** and **PROVOKABLE** should a strategy be in the local environment and ecology? Generosity should be favored in environments (contexts) and ecologies that are both (1) noisy and (2) have a significant proportion of reciprocating strategies and few defectors. When little noise clouds the payments of one’s partners, or when defectors appear plentiful, individuals should prefer strategies that are more **PROVOKABLE**.

²⁵ TFT is easily derived from imitation. It is simply ‘cooperate initially’ and imitate the other guy. Young children readily engage in the kinds of imitative exchanges and behaviors of all kinds (Meltzoff & Prinz, 2002).

Table 1
Strategies by type of partners and context

Categories of partners	Context and ecology	Psychology and behavior
“Close friends”	Substantial noise – exchanges across many domains High b/c Longer memories of important interactions Small number of individuals (memory constraints)	NICE GENEROUS CONTRITE
Distant friends, and other acquaintances	Low noise – in-kind, 1-for-1 exchanges Medium b/c Short memories of interactions Potentially large # individuals	NICE PROVOKABLE
Others	n -person dilemma (public goods situation) Short time horizon (low ω) Low b/c	SUSPICIOUS PROVOKABLE

(3) MEMORY. How much of a partner’s past behavior does one remember and incorporate into her decisions about reciprocation? Overall, individuals in noisy environments should use *selective* forms of longer-term memory (e.g., remember only big defections) for only a small cadre of valued partners. For a large set of partners, memory capacities will be stressed so individuals should prefer strategies that require little memory, but then restrict their interactions to low noise environments and deploy more PROVOKABLE strategies. Ethnographically, we have shown how culture can establish different domains of cooperation in the same social group. In those domains that involve a large number of potential dyadic interactions and a means of reducing the noise in payoffs, cultural evolution favors fairly strict 1-for-1 in-kind exchanges, which make interactions easily trackable and less noisy, and allows short-term accounting for past behavior. In other domains, such as those involving a small number of individuals who make a wide range of different kinds of exchanges (creating lots of noise), favored strategies will involve generosity, contrition, and some kind of longer-term accounting.

Table 1 summarizes the patterns described above. While it seems more likely that the different parameters of reciprocity strategies (e.g., NICENESS) are more accurately thought of as continuous, Table 1’s trichotomy provides a condensed, heuristic presentation of the ideas just discussed.

6.3.9. Indirect reciprocity and reputation

As mentioned above, reciprocity-based approaches to cooperation can take two forms: direct and indirect reciprocity. In *direct reciprocity* favors are exchanged directly and repeatedly between individuals: A helps B and, in return, B helps A. Under *indirect reciprocity*, individuals interact with each other only occasionally (sometimes only once), but now – before interacting – individuals receive information about the past behavior of the individual with whom they are about to interact. For example, A and B

interact and A defects on B (B cooperates); then A and C meet, but because C knows (somehow) about A’s past behavior with B, C defects instead of cooperating (C is normally a cooperator). If C had met B instead of A, C would have cooperated because C knows that B cooperated. C’s knowledge of the behavior of A or B from previous interactions corresponds to what we commonly refer to as reputation. The incorporation of reputation, which provides information about past behavior, generates a plausible mechanism to explain cases of cooperation that involve neither kin nor long-term interactions (Leimar & Hammerstein, 2001; Sugden, 1986; Wedekind & Milinski, 2000). As you will see, the importance of reputation and the practicalities of disseminating reputational information creates a close relationship between two different pathways to cooperation: indirect reciprocity (IR) and social norms and punishment (SNP). In fact, SNP lays a foundation that dramatically strengthens the power of IR.

6.3.10. Indirect reciprocity: solving the core dilemma, again

As discussed above, cooperation can evolve according to the degree by which individuals can preferentially bestow their benefits on cooperators. Theoretical work on indirect reciprocity shows the Core Dilemma once again. This research shows that reciprocating strategies can be maintained against invading defectors when (Nowak & Sigmond, 1998a, 1998b):

$$\phi b > c. \quad (4)$$

As above, b is the benefit of the help contributed, c is the cost of helping to the helper, and ϕ is the probability that the helper accurately knows the reputation of the helpee.

This simple equation, along with extensive theoretical and computer simulation research, shows that the availability of *accurate* reputational information is the key to indirect reciprocity’s ability to solve the puzzle of cooperation (Panchanathan & Boyd, 2003). This means that, *ceteris paribus*, variables such as the size of the cooperative group (the number of individuals in any given interaction), the population size (the number of individuals in the pool of potential interactants), the density of social connections

between individuals in the population, and people's beliefs about gossip will strongly influence the effectiveness of indirect reciprocity. More specifically, the larger the cooperative group, the less a cooperator will be able to direct his help preferentially to individuals with good reputations.²⁶ Likewise, the larger the population the less likely it will be that an individual will have the necessary reputation information about a particular interactant, the more bits of reputational information each individual will have to store in her memory, and the less accurate that information will be. Less dense social networks imply that a cooperator is less likely to have (or be able to get) accurate information about a particular individual. Finally, culturally acquired beliefs about how much to gossip, and how accurate that gossip must be, can greatly influence the effectiveness of indirect reciprocity.²⁷ Individuals, for example, may be punished by others for spreading inappropriate or false gossip. Such punishment is a kind of altruistic punishment in an n -person public goods situation, and is quite difficult to explain without the social norms and punishment (and their supporting psychological mechanisms) discussed below. However, once this punishment of those breaking the gossip norms is in place, much greater amounts of cooperation can be explained by IR.

6.3.11. Cultural capacities vastly increase the potential for indirect reciprocity

Our cultural capacities provide the means to increase both the volume of reputational information available in a population and the accuracy of that information. First, part of our endowment of cultural capacities involves language, and linguistic communication opens up the potential for vast amounts of reputational information to spread around (Smith, 2003). This is only part of the story, as merely increasing the volume of information won't facilitate the evolution of cooperation via indirect reciprocity, *unless* that information remains highly accurate. In this regard, it is important to realize two things: (1) individuals will have incentives to inject false information into the system to either hurt competitors (e.g., spread false rumors about someone), or help themselves (e.g., spread good rumors about themselves), and (2) as information flows from one person to another accuracy will decrease as people misremember, misunderstand and miscommunicate.

²⁶ When a cooperative unit has more than two people (i.e., more than two people are contributing to the provision of a shared good, like taxes or a restaurant bill), it is impossible to withhold cooperation from a defector without also withholding it from cooperators, and this causes all the cooperation in the group to unravel. Here's why: When a cooperator decides not to cooperate in response to a defection, he withholds cooperation not only from the person who previously defected but also from all the people who cooperated. The remaining cooperators then perceive the punisher as a free-rider so they respond by withholding cooperation from him. This process repeats itself until all the reciprocators stop cooperating.

²⁷ See Conte and Paolucci (2002) for an in-depth simulation-based analysis of this issue.

Given all this, how can IR persist at all? Cultural transmission provides a partial rescue.

While not usually thought of in this way, acquiring reputational information from other people about other people is a kind of cultural learning. Consequently, cultural learning mechanisms like prestige-biased transmission and conformist transmission can improve the accuracy of acquiring reputational information for the same reasons that it helps in acquiring adaptive cultural behaviors. This cultural transmission likely involves observing how other individuals behave towards, and what they say about, potential interactants. This includes both what people might intentionally say about someone as well as how they unconsciously behave toward, and speak about, them. For example, prestige-biased transmission allows learners to culturally acquire reputational information about potential interactants from those individuals who are most likely to have good information. For instance, it may involve giving more weight to the opinions individuals who are particularly successful in assessing the reputations of others. Alternatively, it may involve learning from someone who has spent the most time with the individual in question, or from someone who is judged (by the learner) as best able to accurately evaluate the individual in question.

Similarly, conformist transmission allows learners to integrate reputational information received from many people about a particular individual and acquire a more accurate sense of the individual by 'throwing-out' the outliers (this helps remove the noise). For example, suppose people judge the reputation of others on a scale from 0 to 100, and five people have mentally assigned scores of 72, 79, 82, 75 and 0 to "Tom". If a learner samples the words and deeds of these five people with regard to Tom's reputation, our learner might estimate the beliefs of the observed judges as 75, 73, 79, 85, and 0. The learner estimated these scores from observations of the five others. Further suppose that the 'true' reputational score of the individual is 77, and the individual who provided the score of 0 completely misunderstood something Tom did. A conformist learner would throw out the 0 and the 85, and assign the individual a reputation score of 76 (by averaging the remaining 3 estimates). Under a wide range of conditions this allows a better estimate of an individual's 'true reputation' than averaging across all observations equally. Combinations of prestige and conformist mechanisms are a potent means to improve the accuracy of the reputational information that is the lifeblood of Indirect Reciprocity.²⁸ This means that IR will likely *only* be a powerful force in creating cooperation in *highly* cultural species (of which humans are the only one known).

²⁸ Biases like 'self-similarity', sex, age, and ethnicity can further hone the accuracy of one's reputational information (on-average!). For detailed models showing how conformist transmission can improve noisy information, see Henrich and Boyd (2002).

6.3.12. Ethnicity and indirect reciprocity

When individuals encounter someone they don't know and have never heard of (i.e., someone that they cannot easily get reputational information on), what should they do? Theoretical work on IR shows that they should be *SUSPICIOUS*: that is, they should defect and see if the other person cooperates. If the other does cooperate, they should switch to cooperating with that person. This is applicable to a variety of circumstances that might lead some social networks to be more densely interconnected than others. However, vis-à-vis our empirical work and the world more generally, these circumstances often occur when members of an ethnic group tend to preferentially interact mostly with each other, and less with outsiders. Under such conditions, we would expect to see suspicious behavior with non-co-ethnics (because they are outside the reputational network), and nice behavior with co-ethnics. While this observation is important, we do not think it explains the 'ethnic bias' in behavior and interaction. Rather, it's a sociological pattern that arises from individuals interacting with others who carry with them both a psychology for indirect reciprocity and a bias for interaction with co-ethnics (McElreath, Boyd, & Richerson, 2003; Panchanathan & Boyd, 2003). Below we lay out the evolutionary theory that predicts an 'ethnic bias'.

6.3.13. Costly signaling and indirect reciprocity

Are there any circumstances when an evolutionarily successful indirect reciprocator should deliver help to (1) a receiver who is not a known 'cooperator', or (2) a group in which the benefits will be diffused through the group (some of whom may be defectors; this is the *n*-person case)? The answer is 'yes' (Engelmann & Fischbacher, 2002; Panchanathan & Boyd, 2003). This can arise if cooperating (and making a big show of it) can have a sufficiently positive impact on one's reputation in future interaction so as to counteract the immediate costs of helping others (even defectors). This 'reputation building' form of indirect reciprocity corresponds to another class of evolutionary solutions to cooperation termed *costly signaling*.²⁹ In this context, individuals are signaling their quality as cooperators to future interactants using reputational effects. Such cooperative acts are especially useful in situations where many members of a person's social network are present and paying attention. Even if the initial recipient doesn't reciprocate, the giver can have his costs offset by the benefits he receives from the other people who observed his giving and now give to him. Humans should have a psychology that is geared up to look for opportunities of 'high broadcast value' in which

they can obtain big reputational benefits. As we have shown elsewhere (Henrich & Henrich, *in press*), this effect explains a substantial portion of helping behavior in experimental settings and it appears to emerge, intact, in the real world. Among the Chaldeans, for example, men seize the opportunity to pay for everyone's dinner at a restaurant and argue over who will have the chance to pay. The important thing is that everyone knows that they paid. This does not appear to be done with the conscious intent of improving one's reputation; the person truly believes he is being generous solely for the purpose of doing something nice for others – but this does not change the underlying evolutionary basis.

Another implication of this is that individuals should 'cheat the system' when they can by seeking out activities and interactions that enhance their reputation at low cost to themselves. For example, being part of a prosocial charity enhances the reputation of its members. Inside the group, however, free-riding in *n*-person public goods problems reigns because (1) IR is not powerful enough to solve *n*-person dilemmas within the group, and (2) outsiders to the charity don't have access to information about the relative contributions of individual members. Chaldean social life provides a clear case study of this phenomenon.

6.3.14. Psychological and sociological implications

The available theoretical work on indirect reciprocity allows for a series of general predictions about the nature of human psychology and sociology. Psychologically, IR predicts the following:

1. People should care about their reputation, especially when behaving in the presence of members linked to their own social networks.
2. When interacting with individuals not linked to their social network and for whom they have no reputational information, individuals should be *SUSPICIOUS*.
3. Individuals will prefer to be *NICE* to others for whom they lack reputational information if those others are part of their social network. For example, Chaldeans should be *NICE* to other Chaldeans *because* they are tightly connected in dense social networks.
4. Individuals should *prefer* cooperative interactions in dyads or small groups (and will likely not cooperate in larger groups), unless an opportunity presents itself for 'reputation-building'. This will occur mostly in public contexts because they have the most *broadcast value*. Sociologically, IR predicts the following:
5. Dense, bounded social networks that are stable through time lead to the highest levels of IR-based cooperation. Such networks decrease the chance of meeting someone for whom an individual lacks reputational information and increases the opportunity to 'cross-check' information and allow conformist transmission and prestige-bias to sort out the accurate reputational information. Finally, they reduce the proportion of new 'immigrants' who lack sufficient reputational information.

²⁹ We believe that this form of indirect reciprocity is equivalent to forms of costly signaling in which individuals are signaling their value as future cooperative interactants. There are other forms of costly signaling that could lead to cooperative acts that are not part of IR (Bliege Bird, Smith, & Bird, 2001).

6. The most effective forms of cooperation via IR will involve dyads. Public goods problems won't be solved by IR, except under certain conditions when broadcast value is high.
7. Culturally-transmitted beliefs that extend a person's reputation to her kin will facilitate cooperation and increase conformity. Such beliefs have the effect of increasing (sometimes drastically) the impact of having a bad reputation by extending those effects to an individual's close kin, which effectively extends them through time – e.g., a son gets the reputation of his father. Thus, if a father is considering damaging his reputation in some irresistible defection or norm violation, he faces damaging not only himself, but also his son and perhaps his son's son. Here 'cultural evolution' has taken advantage of our kin psychology to enhance the power of IR.

6.4. Social norms, punishment and prosociality

Life in human social groups is regulated by social norms that go beyond cooperation. Norms can be identified by three characteristics (Henrich et al., 2003): (1) they prescribe 'proper' behavior for individuals within a population, or some subset of the population (e.g., women must wear veils), (2) these prescriptions are widely shared by at least some important portion of the population, and (3) failure to adhere sufficiently closely to these prescriptions will anger other individuals in the population (even if the action does not otherwise affect them), and these angry individuals may take actions that are costly to both themselves and the norm violator(s). Some examples from anthropology and common experience include eating tabooed food (e.g., pork), going nude at formal weddings, having sex with one's parents, not cutting your lawn frequently enough, telling jokes during a funeral services, etc. None of the above theories can explain these pan-human patterns. Why would uninvolved third parties care about what others do, and why would they care enough to take an action that is costly to themselves? And what does this 'norm stuff' have to do with human cooperation?

In thinking about norms, it is important to start with individual minds and then aggregate up (using mathematical models of social interaction and learning) to population-level phenomena. At the individual level, norms begin as sets of mental representations, stored in individual's brains, which we commonly refer to as ideas, preferences, beliefs, values and practices. These representations prescribe both what their possessor should do in certain situations, and what others should do. As prescriptions, norms are often attached to powerful emotions (anger, guilt and shame) and motivations that lead to strong reactions when the individual, or others, violate a norm. These reactions may lead to a variety of forms of punishment, ranging from gossip to banishment to homicide.

At the population-level, these mental representations (norms) are shared to some degree, for a variety of reasons,

by many members of the same social group. One important reason is that cultural learning mechanisms (like conformist transmission) will cause members of a social group to adopt similar mental representations. Second, the punishment evoked by norms will further lead individuals to adopt practices (if not similar mental representations) to avoid punishment. There are other reasons, but these two classes of mechanisms alone will cause groups to share similar mental representations (norms) about various kinds of behavior. *Some* of these representations and behaviors have to do with cooperation.

Social norms, despite being one of the most discussed concepts in the social sciences (Bendor & Swistak, 2001), have lacked a serious evolutionary explanation that can account for their character and diversity. We think a firm evolutionary explanation for social norms lies in the fact that humans – unlike other animals – rely heavily on their evolved cognitive adaptations for cultural learning to acquire a large portion of their behavioral repertoire, including their social behavior. When both adherence to a norm and a willingness to punish norm violations are influenced by cultural learning, the mechanisms of prestige-biased and conformist transmission can lead to stable situations in which most people acquire and follow the rules, prescriptions and punishments associated with a social norm. This applies to any norm, be it adaptive, neutral or maladaptive, and includes norms for costly cooperation. This body of formal theoretical work further suggests that, while neutral and somewhat maladaptive norms could be maintained *within* any particular group, group beneficial norms can spread by competition and selection *across* social groups that have different culturally evolved norms that vary in their group-beneficial properties – a process termed *cultural group selection*. Furthermore, if these competitions among groups with different norms have been occurring for a long time (tens of thousands of years), the theory shows that the punishment of norm violators within a group will cause natural selection to favor 'prosocial genes' (genes that would favor Darwin's "high standards of morality", see the quotation opening this paper).³⁰ As a term of reference, we will call the evolved aspects of human psychology derived from this kind of culture-gene coevolutionary process our *Prosocial Psychology* (Henrich, 2004).

6.4.1. Cultural learning and punishment leads to social norms

The interaction of cultural learning and punishment leads to norms that are locally stable in social groups. Although the mathematics showing that a combination of prestige-biased transmission and conformist transmission will lead to stable cooperative norms is somewhat

³⁰ Some students of evolution and human behavior have a visceral reaction to anything that uses the term "group selection". Fear not. Cultural group selection models do NOT carry the problems typically associated with models of genetic group selection (Boyd, Gintis, Bowles, & Richerson, 2003; Henrich, 2004; Smith & Bliege Bird, 2000).

more complicated than that used for models of kinship and reciprocity (Henrich & Boyd, 2001), the logic is fairly straightforward. Prestige-bias transmission generally acts to favor the imitation of behaviors, values and practices that lead to the most successful (and often the most fitness enhancing) behaviors. Conformist transmission acts to favor the spread of the majority's behavior/practices, and principally acts when the difference in the success between various behaviors or strategies is small or difficult to figure out. That is, conformist transmission can only maintain behaviors in a group if they are neutral, not too costly, and/or sufficiently ambiguous. Thus, for most important cases of cooperation, or other costly norms, we expect that the costs of doing a cooperative behavior (the 'norm') will be too costly for conformist transmission to have any important effects.³¹

Adding culturally-transmitted tendencies or 'tastes' for punishing norm violators can turn the tables on this logic. Imagine a group in which strategies for both punishing norm violators and adhering to a costly norm-behavior (e.g., cooperation) are common. Punishing norm violators is costly compared to not punishing norm violators, so many learning and decision-making mechanisms will still favor not-punishing. However, punishing norm violators can be substantially *less costly* than adhering to a norm itself, so punishment of norms is easier to stabilize than many costly norms. Here's why: If our group consists mostly of individuals who adhere to the norm and punish those who do not, then most people will stick to the norm (e.g., recycle tin cans) in order to avoid the costs of being punished. If punishers are common, the cost of being punished can easily exceed the costs of sticking to a costly norm. But, if everyone sticks to the norm (because of punishment), punishers don't have to do anything, and being a punisher is not particularly costly. If some few individuals occasionally violate the norm (say, by mistake), punishing costs can still remain pretty small. Because these punishing costs are substantially smaller than those associated with the cost of sticking to the norm itself, conformist transmission can often maintain the punishment of norm violators in the population, against forces like prestige-bias transmission and experiential learning, even when conformist transmission is not strong enough to stabilize the costly norm itself (without any punishment). How well this trick works depends on the ratio of the costs of punishing to the costs of being punished. The more effective punishment is (i.e., the smaller the ratio of the costs of punishing to the

costs of being punished), the more costly can be the norms that are maintained. In summary, this can work because (1) punishers don't have to pay the costs of punishing very often if being punished is more costly than the costs associated with sticking to the norm – everyone generally sticks to the norm, and punishers need only punish occasional deviants; (2) the cost of punishing is small (and probably ambiguous), so conformist transmission can overcome it, and keep a strategy of punishing stable in the social group; and (3) when punishing norm violators is common, everyone tends to adhere to the norm because the costs of being punished for violating the norm exceed the costs of sticking with the norm.³²

However, suppose that conformist transmission is not strong enough to stabilize the punishment of norm violators. Can costly norms still be maintained? Under such circumstances it is possible that conformist transmission can act on a tendency to punish individuals who fail to punish norm violators. We call this the 'punishment of non-punishers'. By the same logic as above, if a group contains mostly individuals who adhere to a norm, punish violators of the norm, and punish those who fail to punish norm violators, conformist transmission can stabilize all of these behaviors by maintaining the least costly behavior – the punishment of those who fail to punish norm violators. This behavior is the 'cheapest' way to maintain the norm because those who punish those who fail to punish norm violators will only have to pay a cost when a norm violation occurs (which are rare to begin with) and someone fails to punish that violation (another rare event). This means that these costs of punishing need only be paid after the conjunction of two rare events – making it an order of magnitude less costly than punishing norm violators. By favoring the punishment of non-punishers, the strategy of punishing norm violators is maintained because the costs of being punished for not punishing norm violators exceeds the cost of occasionally administering punishment, and in turn, the presence of those who punish norm violators means the costs of being punished for violating the norm can exceed the costs of sticking to the norm. In this way, conformist transmission can indirectly stabilize quite costly norms.

Actually, the mathematical analysis behind the above idea is even more nuanced. It shows that 'b' (the benefits from the cooperative act—see above), does *not* influence the creation of a stable norm! That is, the math shows that a combination of prestige and conformist biases and punishing behaviors can stabilize any costly norm ('c', the cost of the behavior, does matter) independent of whether it benefits anyone. If conformist transmission can favor the punishment of a behavior then punishment will cause the behavior, practice or belief to stay common in the group.

³¹ In analyzing the evolution of cooperation, the real challenge is to explain the maintenance or stability of cooperative and prosocial (punishment of non-cooperators) behavior or tendencies, not their spread from zero. All kinds of random effects, population shocks, stochastic migrations patterns, etc. to the evolving system can make cooperators common at a particular place and time, but the key is to explain why these cooperators won't simply vanish back into evolutionary history as the system returns to equilibrium. The same situation holds true for reciprocity solutions; TFT, for example, is not favored until it is sufficiently common.

³² Boyd et al. (2003) take advantage of this logic in a computer simulation to show that, even without conformist transmission, the introduction of punishing strategies substantially enhances the effect of competition between social groups, and thereby favors the evolution of substantial amounts of pro-group cooperation.

The benefit, b , of the behavior could be zero, or *negative*, and the group could still ‘lock in’ on the norm. This was an unexpected prediction that fell out of theoretical work on cooperation (Boyd & Richerson, 1992), and just happens to explain the massive database of anthropological findings showing that many idiosyncratic social norms are either neutral or even maladaptive (Edgerton, 1992). While this aspect of the theory explains social norms in general (and cooperative norms are one kind of costly norm), it does not explain why cooperative social norms should be more common, or more likely to spread, than other kinds of norms.

The answer to “why are there so many cooperative social norms?” is explained by *cultural group selection* (CGS). As different social groups arrive at and ‘lock in’ on different social norms, CGS provides a process for selection among alternative social norms. Some groups will develop norms about constructing community buildings, not eating snakes and fishing in cooperative units, while other groups may culturally evolve norms about cooperatively raiding other groups, sending children to school, and giving young girls cliterectomies. With these different norms in place, social groups can compete in a variety of ways. First, some groups have cooperative norms that yield greater success in warfare, and they may spread their norms by conquering other groups (Soltis, Boyd, & Richerson, 1995). Second, some groups have norms (probably cooperative) that increase their economic (making more food or tools) or demographic (more kids) production, so they may spread their norms by generating more carriers of the norm than other groups. For example, if a group has economic practices that enable them to better feed their children than neighboring groups, then they will likely have more children that survive to spread their norms. This effect would be magnified if the group also had sexual norms that led to higher reproductive rates than other groups. Or third, if members of different social groups have sufficient interaction (which they often do), prestige-biased transmission can lead individuals to preferentially imitate people from more successful groups, such as groups with higher qualities of life in terms of health, housing and material possessions. Thus, the norms of a successful group can preferentially spread from group to group relatively rapidly (Boyd & Richerson, 2002a). Each of the above processes can lead to the preferential proliferation of cooperative norms, and each has been observed in the ethnographic, archaeological and historical record (Atran et al., 2002; Diamond, 1997; Flannery & Marcus, 2000; Kelly, 1985; Shennan, 2003; Soltis et al., 1995; Stark, 1997; Wilson, 2002).

One of the best-documented cases of cultural evolution through intergroup competition occurred during the 18th century among the anthropologically famous ethnic groups of the Nuer and Dinka. Before 1820, the Nuer and Dinka (Kelly, 1985) occupied adjacent regions in the southern Sudan. Although the groups inhabited similar environments and possessed identical technology, they differed in significant ways. Economically, both the Dinka and Nuer raised cattle, but the Dinka maintained smaller herds of

approximately nine cows per bull, while the Nuer maintained larger herds with two cows per bull. The Nuer ate mostly milk, corn, and millet and rarely slaughtered cows, while the Dinka frequently ate beef. Politically, the Dinka lived in small groups, the largest of which corresponded to their wet season encampment. In contrast, the Nuer organized according to a patrilineal kin system that structured tribal membership across much larger geographic areas. Consequently, the size of a Dinka social group was limited by geography, whereas the Nuer system could organize much larger numbers of people over greater expanses of territory. Despite the similarity of their environments, these two groups showed substantial economic and political differences. Over about 100 years, starting around 1820, the Nuer dramatically expanded their territory at the expense of the Dinka, who were driven off, killed, or captured and assimilated. As a result, Nuer beliefs and practices spread fairly rapidly across the landscape relative to Dinka beliefs and practices – even though the Nuer were soon living in the once “Dinka environment” and the fact that many Nuer were formerly Dinka who had adopted Nuer customs (so the cause of the differential success can’t be attributed to environmental or genetic differences).

6.4.2. *Genes respond in the wake of cultural group selection*

Imagine yourself back in human prehistory, say 55,000 years ago. Social groups are competing and cooperative norms of various types are spreading through the human species via cultural learning. Groups have different sets of norms, including those that promote cooperative hunting, the sharing of tool-making techniques, community house building, trade (which requires norms for interaction), raiding, and warfare. The spreading of these different norms effectively changes the ‘selective environment’ faced by genes, because now successful genes have to adapt an individual to a world in which one is punished for norm violations – and norms are often cooperative or prosocial. This culturally-evolved selection pressure will favor genes that allow individuals to rapidly acquire the local norms (thereby avoiding punishment), and avoid the temptation of norm violation (i.e., defection). Once the newly selected for genes have spread, cultural learning mechanisms will favor even larger, more costly cooperative norms, and CGS will continue to spread those norms that allow social groups to compete more effectively with other groups (Henrich & Boyd, 2001). Gradually this interaction of cultural and genetic transmission ratcheted up our sociality, honed our preferences for helping others, led to punishing violators and avoiding punishment, hardened our ability to inhibit quick defections, and refined our learning capacities to promote the efficient acquisition of norms until we became the only ultra-social primate (Richerson & Boyd, 1998). The Prosocial Psychology that underpins much of our contemporary social life had emerged.

The theory just described above, leads to several general predictions about the nature of human sociality and our social psychology:

1. Different human groups will be characterized by different social norms – some cooperative, some not. Some norms will be maladaptive. Non-cultural species will not show this kind of variability.
2. Negative reactions by third parties to norm violations will be a human universal (third parties are individuals who suffer little or no cost from the norm violation). Non-cultural species will not show third-party punishment of this kind.
3. More costly norms, such as cooperative norms, will involve costly punishment of norm violators by third parties, and possibly the punishment of non-punishers. This also implies that *very* low cost cooperative behaviors can be favored by conformist transmission alone, without punishment.
4. Human social groups will vary in their ‘willingness to punish’, and this willingness to engage in costly punishment will be context specific. When applied to cooperative interactions this ‘willingness to punish’ accounts for the empirical phenomena of *altruistic punishment* and third-party punishment that has been rigorously documented in experimental work (Fehr & Fischbacher, 2003; Fehr, Fischbacher, & Gaechter, 2002; Fehr & Gächter, 2002).
5. Norms will usually be context specific. Using experimental data, we and others have shown how context matters (Henrich & Henrich, *in press*).
6. The strength of norms depends on the ratio of the cost of punishing another individual to the cost of being punished. Consequently, social situations that allow for effective punishment to be dealt out at low cost will favor the maintenance of group norms, including cooperative norms.
7. In groups with dense social networks, punishment can operate through damaging of reputation. This provides a cheap and effective means to punish norm violators.
8. Cultural beliefs that extend the impact of punishment (such as by effecting the reputation of close kin of the punished person) increase the effectiveness of punishment for norm maintenance. This parallels our earlier discussion of reputation and indirect reciprocity.
9. Social norms that facilitate the spreading of accurate reputational information, in part by punishing those who spread false reputational information, can lay a foundation for substantially increasing the degree of cooperation that can be supported by Indirect Reciprocity.

6.5. Ethnicity, norms and cooperation

The evolution of norms lays the ground work for the culture-gene coevolution of an ‘ethnic psychology’ and the sociological emergence of ‘ethnic groups’. Our theorizing begins by considering how the cultural evolution of norms yields new selection pressures on genes. Two problems present themselves. First, individuals have to figure out what the ‘right norm’ is for getting along in their social groups, keeping in mind that different social groups evolve different norms. By ‘right’ we mean the norm that allows the individ-

ual to avoid punishment for norm violations and coordinate their behavior with other members of their social group. Because norms can have an arbitrary character, natural selection will favor genes that direct individuals to learn from those people who are most likely to have the ‘right norms’. The second problem arises because, once an individual has adopted some norms, he would be best off to interact with other individuals who also share his norms – otherwise he may get punished and/or miscoordinate with those with whom he is interacting. For example, if a guy believes that his wife should come with a dowry, he needs to find women whose families believe that they should pay a dowry. The root of the puzzle is that people’s norms are not stamped on their foreheads. If people knew the underlying norms of all people, they could be sure to interact with those sharing their norms. The tricky part is that underlying norms are more typically hidden properties of individuals that rarely surface, and only in specific situations.

So far we have focused on norms enforced by punishment, especially ‘cooperative norms’. However, the evolution of our ethnic psychology and ethnic groups can be equally well explained by ‘coordination norms’. Unlike cooperative norms, situations involving coordination problems do not have a free rider problem. Individuals achieve the highest payoff or fitness by doing what everyone else is doing. For example, the decision to drive on the right or left side of the road is a coordination problem. If you are in England, you want to drive on the left, and if you are in Germany it’s best to drive on the right. Human societies are full of coordination problems, but many of them are substantially more *moralized* than ‘which side to drive on’. Consider the marriage customs of some social groups, which demand that a dowry be sent along with their daughter to the groom, or to the family of the groom. In others groups, the groom or the groom’s family (or both) are expected to pay for the bride in, for example, cows, service (labor to the bride’s family) or precious metals. Miscoordinations occur when both bride and groom’s families are expecting payment from the other, and usually everyone loses as no wedding occurs. In a sense, conformist transmission and punishment turn cooperative dilemmas into these coordination situations, so both can contribute to the emergence of an ethnic psychology and ethnic groups.

Now back to the two problems: How does natural selection: (1) figure out who to learn norms from so as to avoid getting the ‘wrong’ norms, and (2) make sure individuals preferentially interact with others most likely to share their norms? Recent theoretical work on the question (McElreath et al., 2003) suggests that selection ‘looks’ for statistically reliable correlations between ‘observables’ and these underlying norms.³³ Natural selection is fortunate in this

³³ Note the parallel with our discussion of kin selection, where we explained how ‘smell’ or ‘hanging around mom’ can provide reliable predictors of relatedness, and thus reliable predictors on ‘having the same gene’.

case because the same cultural learning processes that transmit the behaviors and beliefs related to social norms also often transmit things like dress, language, accent, dialect, behavioral mannerisms and food preferences. Therefore, this theoretical work predicts that human psychology evolved to seek out ‘indicator traits’ (language, dress, etc.) that match its own because such markers tend to reliably co-occur with the ‘right norms’ – ‘right’ meaning both the norms a learner wants to learn, and the norms that match one’s own. Using such markers, individuals can bias both their learning and interaction towards those individuals who share their same culturally-transmitted ‘indicator traits’. As above, the proximate psychological mechanism of this ‘ethnic psychology’ will involve attentional biases (who’s interesting?) and affective motives (who does one tend to ‘like’ being around?).

Building on this theory, new empirical work suggests that people tend to think about certain human groups (those that we’d typically call ‘ethnic groups’) in the same way that humans think about different biological species: we understand the characteristics and attributes of these groups and biological species in both *essentialist* (all members carry the same unchangeable essence that gives rise to their shared characteristics) and *primordialist* (this ‘essence’ is transmitted down blood lines) terms (Gil-White, 1999, 2001). In the case of ethnic groups, this empirical work suggests that people see ethnic markers or ethnic membership as a cue of hidden, underlying qualities or fundamental (unchangeable) attributes. The three characteristics of a group that tend to spark this type of ‘species thinking’ are: (1) shared markings (shared language, dress, etc.), (2) similarity between parents and offspring, and (3) group endogamy (marriage and mating occurs within the ‘ethnic’ group) and common descent. The greater the degree to which these characteristics are present in a human group – or the greater the degree to which individuals believe or perceive these as present, the more likely individuals are to use ‘species thinking’ vis-à-vis members of that group. The more individuals use these modes of essentialist and primordialist thinking, the stronger the ethnic bias in interaction and learning, and the greater the importance of ethnic membership on social behavior and cooperation.³⁴

³⁴ We’ve not the space to deal with this at any length, but this theory of ethnicity explains the essentialist and primordial nature of human thinking vis-à-vis race as a byproduct of the coevolutionary processes that produced it for ethnicity (Gil-White, 1999, 2001). In human ancestral history, individuals probably rarely, if ever, encountered individuals who looked very different from themselves (i.e., biologically: skin color, morphology, hair, etc.). When people did finally meet other such human groups late in human history, these phenotypic differences, by chance, fit easily with the cues for different ethnic groups – racial difference may super-stimulate this tendency to ‘species-thinking’. To be perfectly clear, we are NOT saying that there are essential or primordial differences between either racial or ethnic groups. We are addressing why humans so frequently (and incorrectly) think about such groups in these terms.

Understanding human ethnic-psychology, and the coevolutionary processes that lie behind it, illuminates a large number of puzzling patterns in the world. Perhaps the most general puzzle is why ethnically-marked groups seem so important in the world in comparison to other kinds of human groups: Why are so-called ethnic boundaries so important in the world compared to other possible boundaries? Why are political parties so ethnically biased (instead of class or ‘common interest’ biased)? Why do people often support political candidates that share their ethnic markers? Why does violence, oppression and warfare so often fall along ethnic lines? Why is ethnicity so important in marriage and sex? The theory summarized above leads to a series of predictions that help address some of these puzzles:

1. People (children and adults) use ethnic cues to figure out from whom to learn. This kind of learning involves biases in both attention and memory.
2. People prefer to interact with individuals who share their ethnic markers.
3. At a sociological level, these psychological learning biases cause individuals who share ethnic markers to share lots of other norms, beliefs and values.
4. This psychological preference also creates all kinds of ‘ethnic clumping’, as people seek out members of their own ethnic groups in marriage, clubs, religion, politics, etc.
5. Ethnic markers tend to be ‘hard-to-fake’ (honest signals), as these provide the most reliable cues of the underlying norms. Language and dialect are particularly important, as these cannot be easily learned and can rarely be perfectly faked.

Elsewhere we summarize some of the field and experimental evidence for these empirical entailments (Henrich & Henrich, *in press*, chap. 8).

7. Conclusion

Our goal here has been to integrate an understanding of cultural evolution into a broader culture-gene coevolutionary framework. As we see it, one of the main problems with standard evolutionary approaches is that they fail to account for the fact that humans, to a degree not observed in other animals, have evolved to rely heavily on observational learning to acquire vast swaths of their behavioral repertoire – a process that can, in turn, influence the direction of genetic evolution. On the other hand, traditional ‘cultural approaches’ tend to ignore or deny any influence of evolution on human behavior, and seem blind to the fact the cultural learning itself is best understood as a kind of adaptation, whose origins and operation can be rigorously examined with a combination of formal models, ethnographic inquiry, and focused experimentation.

References

- Alvard, M. (2003). The adaptive nature of culture. *Evolutionary Anthropology*, 12(3), 136–147.
- Atran, S., Medin, D., Ross, N., Lynch, E., Vapnarsky, V., Ek', E. U., et al. (2002). Folkeology, cultural epidemiology, and the spirit of the commons: a garden experiment in the Maya lowlands, 1991–2001. *Current Anthropology*, 43(3), 421–450.
- Atran, S., Medin, D., Ross, N., Lynch, E., Coley, J., Ek', E. U., & Vapnarsky, V. (1999). Folkeology and commons management in the Maya lowlands. *Proceedings of the National Academy of Sciences*, 96, 7598–7603.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Axelrod, R., & D'Ambrosio, L. (1994). Annotated bibliography on the evolution of cooperation. Unpublished manuscript, Ann Arbor.
- Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, 211, 1390–1396.
- Baldwin, J. M. (1896a). A new factor in evolution. *American Naturalist*, 30(441–457), 536–554.
- Baldwin, J. M. (1896b). Physical and social heredity. *American Naturalist*, 30, 422–428.
- Bendor, J. (1987). In good times and bad: reciprocity in an uncertain world. *American Journal of Political Science*, 31, 531–558.
- Bendor, J. (1993). Uncertainty and the evolution of cooperation. *Journal of Conflict Resolution*, 37(4), 709–734.
- Bendor, J., Kramer, R. M., & Stout, S. (1991). When in doubt ... Cooperation in a noisy prisoners-dilemma. *Journal of Conflict Resolution*, 35(4), 691–719.
- Bendor, J., Kramer, R., & Swistak, P. (1996). Cooperation under uncertainty: what is new, what is true, and what is important? *American Sociological Review*, 61(2), 333–338.
- Bendor, J., & Mookherjee, D. (1987). Institutional structure and the logic of ongoing collective action. *The American Political Science Review*, 81(1), 129–154.
- Bendor, J., & Swistak, P. (1997). The evolutionary stability of cooperation. *American Political Science Review*, 91(2), 290–307.
- Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, 107(2), 1493–1545.
- Bliege Bird, R. L., Smith, E. A., & Bird, D. W. (2001). Bliege the hunting handicap: costly signaling in human foraging societies. *Behavioral Ecology and Sociobiology*, 50, 9–19.
- Boerlijst, M. C., Nowak, M. A., & Sigmund, K. (1997). The logic of contrition. *Journal of Theoretical Biology*, 185, 281–293.
- Boyd, R. (1989). Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *Journal of Theoretical Biology*, 136, 47–56.
- Boyd, R. (1992). The evolution of reciprocity when conditions vary. In F. Dewaal & A. H. Harcourt (Eds.), *Coalitions in humans and other animals* (pp. 473–492). New York: Oxford University Press.
- Boyd, R., Gintis, H., Bowles, S., & Richerson, P. J. (2003). The evolution of altruistic punishment. *PNAS*, 100(6), 3531–3535.
- Boyd, R., & Lorderbaum, J. P. (1987). No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature*, 32(6117), 58–59.
- Boyd, R., & Richerson, P. (1988). The evolution of reciprocity in sizable groups. *Journal of Theoretical Biology*, 132, 337–356.
- Boyd, R., & Richerson, P. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology & Sociobiology*, 13(3), 171–195.
- Boyd, R., & Richerson, P. (2002a). Group beneficial norms can spread rapidly in a structured population. *Journal of Theoretical Biology*, 215, 287–296.
- Boyd, R., & Richerson, P. (2002b). Solving the puzzle of human cooperation. In S. Levinson (Ed.), *Evolution and culture*. Cambridge MA: MIT Press.
- Boyd, R., & Richerson, P. J. (1985). *Culture and the evolutionary process*. Chicago, IL: University of Chicago Press.
- Boyer, P. (2001). *Religion explained: The evolutionary origins of religious thought*. New York: Basic Books.
- Chagnon, N., & Irons, W. (1979). *Evolutionary biology and human social behavior: An anthropological perspective*. North Scituate: Duxbury Press.
- Conte, R., & Paolucci, M. (2002). *Reputation in artificial societies: Social beliefs for social order*. Boston: Kluwer Academic Publishers.
- Cronk, L., Chagnon, N., & Irons, W. (Eds.). (2000). *Adaptation and human behavior*. New York: Aldine De Gruyter.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- DeBruine, L. (2002). Facial resemblance enhances trust. *Proceedings of the Royal Society of London Series B – Biological Sciences*, 269, 1307–1312.
- Diamond, J. M. (1997). *Guns, germs, and steel: The fates of human societies*. New York: W.W. Norton & Co.
- Durham, W. H. (1991). *Coevolution: Genes, culture, and human diversity*. Stanford, CA: Stanford University Press.
- Durkheim, E. (1933). *The division of labor in society*. New York: Free Press.
- Edgerton, R. B. (1992). *Sick societies: Challenging the myth of primitive harmony*. New York: Free Press.
- Engelmann, D., & Fischbacher, U. (2002). Indirect reciprocity and strategic reputation building in an experimental helping game. Unpublished manuscript, University of Zurich.
- Farrell, J., & Ware, R. (1989). Evolutionary stability in the repeated prisoner's dilemma. *Theoretical Population Biology*, 36(161–166).
- Fehr, E., & Fischbacher, U. (2003). Third party sanction and social norms. *Evolution and Human Behavior*, 25(2004), 63–87.
- Fehr, E., Fischbacher, U., & Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nature*, 13, 1–25.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415, 137–140.
- Fehr, E., & Henrich, J. (2003). Is strong reciprocity a maladaptation? In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.
- Fiske, A. (1991). *Structures of social life*. New York: Free Press.
- Flannery, K. V., & Marcus, J. (2000). Formative Mexican chiefdoms and the myth of the 'mother culture'. *Journal of Anthropological Archaeology*, 19(1–37).
- Frank, S. (1995). George price's contributions to evolutionary genetics. *Journal of Theoretical Biology*, 175, 373–388.
- Frank, S. (1997). The price equation, Fisher's fundamental theorem, kin selection, and causal analysis. *The Society for the Study of Evolution*, 51(6), 1712–1729.
- Frank, S. (1998). *Foundations of social evolution*. Princeton: Princeton University Press.
- Gil-White, F. (1999). How thick is blood? The plot thickens. *Ethnic and Racial Studies*, 22(5), 789–820.
- Gil-White, F. (2001). Are ethnic groups biological 'species' to the human brain? Essentialism in our cognition of some social categories. *Cultural Anthropology*, 42(4), 515–554.
- Grafen, A. (1985). Hamilton's rule OK. *Nature*, 318, 310–311.
- Hamilton, W. D. (1964). The genetical evolution of social behavior. *Journal of Theoretical Biology*, 7, 1–52.
- Hamilton, W. D. (1972). Altruism and related phenomena, mainly in the social insects. *Annual Review of Ecology and Systematics*, 3, 192–232.
- Hammerstein, P. (2003). Why is reciprocity so rare in social animals? In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.
- Hawkes, K., O'Connell, J. F., & Blurton Jones, N. (2001). Hadza meat sharing. *Evolution and Human Behavior*, 22, 113–142.
- Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53, 3–35.
- Henrich, J. (in press). A cultural species. In M. Brown (Ed.), *Explaining culture scientifically*. Seattle: University of Washington Press.

- Henrich, J., Bowles, S., Smith, E. A., Young, P., Boyd, R., Sigmund, K., et al. (2003). The cultural and genetic evolution of human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.
- Henrich, J., & Boyd, R. (1998). The evolution of conformist transmission and the emergence of between-group differences. *Evolution and Human Behavior, 19*, 215–242.
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of Theoretical Biology, 208*, 79–89.
- Henrich, J., & Boyd, R. (2002). On modeling cultural evolution: why replicators are not necessary for cultural evolution? *Journal of Cognition and Culture, 2*(2), 87–112.
- Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., & Gintis, H. (Eds.). (2004). *Foundations of human sociality: Economic experiments and ethnographic evidence from fifteen small-scale societies*. Oxford: Oxford University Press.
- Henrich, J., & Gil-White, F. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior, 22*(3), 165–196.
- Henrich, J., & McElreath, R. (2003). The evolution of cultural evolution. *Evolutionary Anthropology, 12*(3), 123–135.
- Henrich, N. S., & Henrich, J. (in press). *Why humans cooperate: A cultural and evolutionary explanation*. Oxford: Oxford University Press.
- Hill, K. (2002). Altruistic cooperation during foraging by the ache and the evolved human predisposition to cooperate. *Human Nature, 13*(1), 105–128.
- Hirshleifer, J., & Martinez-Coll, J. C. (1988). What strategies can support the evolutionary emergence of cooperation? *Journal of Conflict Resolution, 32*(2), 367–398.
- Johnson, A., & Earle, T. (2000). *The evolution of human societies* (2nd ed.). Stanford: Stanford University Press.
- Joshi, N. V. (1987). Evolution of cooperation by reciprocation within structured demes. *Journal of Genetics, 66*(1), 69–84.
- Kameda, T., & Nakanishi, D. (2002). Cost-benefit analysis of social/cultural learning in a non-stationary uncertain environment: an evolutionary simulation and an experiment with human subjects. *Evolution and Human Behavior, 23*, 373–393.
- Keller, L., & Ross, K. G. (1998). Selfish genes, a green beard in the red fire ant. *Nature, 394*, 573–575.
- Kelly, R. C. (1985). *The Nuer conquest*. Ann Arbor: University of Michigan Press.
- Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings: Biological Sciences, 268*(1468), 745–753.
- Lorberbaum, J. (1994). No strategy is evolutionarily stable in the repeated prisoner's dilemma. *Journal of Theoretical Biology, 168*, 117–130.
- Lorberbaum, J. P., Bohning, D. E., Shastri, A., & Sine, L. (2002). Are there really no evolutionary stable strategies in the iterated prisoner's dilemma? *Journal of Theoretical Biology, 214*, 155–169.
- Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge: Cambridge University Press.
- McElreath, R., Boyd, R., & Richerson, P. (2003). Shared norms can lead to the evolution of ethnic markers. *Current Anthropology, 44*(1), 123–129.
- Meltzoff, A. N., & Prinz, W. (Eds.). (2002). *The imitative mind*. New York: Cambridge University Press.
- Nowak, M. A., & Sigmund, K. (1998a). The dynamics of indirect reciprocity. *Journal of Theoretical Biology, 194*, 561–574.
- Nowak, M. A., & Sigmund, K. (1998b). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573–577.
- Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: the importance of standing for the evolution of indirect reciprocity. *Journal of Theoretical Biology, 224*, 115–126.
- Platak, S. M., Burch, R. L., Panyavin, I. S., Wasserman, B. H., & Gallup, G. G. (2002). Reactions to children's faces: resemblance affects males more than females. *Evolution and Human Behavior, 23*, 159–166.
- Platak, S. M., Critton, S. R., Burch, R. L., Frederick, D. A., Myers, T. E., & Gallup, G. G. (2003). How much paternal resemblance is enough? Sex differences in hypothetical investment decisions but not in the detection of resemblance. *Evolution and Human Behavior, 24*, 81–87.
- Queller, D. C. (1992). A general model for kin selection. *Evolution, 46*(2), 376–380.
- Richerson, P., & Boyd, R. (1998). The evolution of ultrasociality. In I. Eibl-Eibesfeldt & F. K. Salter (Eds.), *Indoctrinability, ideology and warfare* (pp. 71–96). New York: Berghahn Books.
- Richerson, P., & Boyd, R. (2000). Complex societies: the evolutionary dynamics of a crude superorganism. *Human Nature, 10*, 253–289.
- Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution*. Chicago: University of Chicago Press.
- Rogers, E. M. (1995). *Diffusion of innovations*. New York: Free Press.
- Selten, R., & Apeste-gula, J. (2002). Experimentally observed imitation and cooperation in price competition on the circle. Unpublished manuscript.
- Shennan, S. (2003). *Genes, memes, and human history: Darwinian archaeology and cultural evolution*. London: Thames & Hudson.
- Smith, A. (2000). *The theory of moral sentiments*. New York: Prometheus Books.
- Smith, E. A. (2003). Human cooperation: perspectives from behavioral ecology. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation*. Cambridge: MIT Press.
- Smith, E. A., & Bliege Bird, R. L. (2000). Turtle hunting and tombstone opening: public generosity as costly signaling. *Evolution and Human Behavior, 21*(4), 245–261.
- Soltis, J., Boyd, R., & Richerson, P. J. (1995). Can group-functional behaviors evolve by cultural group selection? An empirical test. *Current Anthropology, 36*(3), 473–494.
- Stark, R. (1997). *The rise of christianity: How the obscure, marginal Jesus movement became the dominant religious force in the Western World in a few centuries*. New York: Harper Collins.
- Sugden, R. (1986). *The economics of rights, co-operation, and welfare*. Oxford: Basil and Blackwell.
- Tomasello, M. (1999). The human adaptation for culture. *Annual Reviews of Anthropology, 28*, 509–529.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46*, 35–57.
- Trivers, R. (1985). *Social evolution*. Menlo Park, CA: Benjamin/Cummings.
- Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science, 288*, 850–852.
- Wilson, D. S. (2002). *Darwin's Cathedral: Evolution, religion, and the nature of society*. Chicago: University of Chicago.
- Wolf, A. P. (1995). Sexual attraction and childhood association: A Chinese brief for Edward Westermarck. *Stanford University Press*.
- Wu, J., & Axelrod, R. (1995). How to cope with noise in the iterated prisoner's dilemma. *Journal of Conflict Resolution, 39*(1), 183–189.