

Feature Review

Human cooperation

David G. Rand¹ and Martin A. Nowak²¹ Department of Psychology, Department of Economics, Program in Cognitive Science, School of Management, Yale University, New Haven, CT, USA² Program for Evolutionary Dynamics, Department of Mathematics, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

Why should you help a competitor? Why should you contribute to the public good if free riders reap the benefits of your generosity? Cooperation in a competitive world is a conundrum. Natural selection opposes the evolution of cooperation unless specific mechanisms are at work. Five such mechanisms have been proposed: direct reciprocity, indirect reciprocity, spatial selection, multilevel selection, and kin selection. Here we discuss empirical evidence from laboratory experiments and field studies of human interactions for each mechanism. We also consider cooperation in one-shot, anonymous interactions for which no mechanisms are apparent. We argue that this behavior reflects the overgeneralization of cooperative strategies learned in the context of direct and indirect reciprocity: we show that automatic, intuitive responses favor cooperative strategies that reciprocate.

The challenge of cooperation

In a cooperative (or social) dilemma, there is tension between what is good for the individual and what is good for the population. The population does best if individuals cooperate, but for each individual there is a temptation to defect. A simple definition of cooperation is that one individual pays a cost for another to receive a benefit. Cost and benefit are measured in terms of reproductive success, where reproduction can be cultural or genetic. **Box 1** provides a more detailed definition based on game theory. Among cooperative dilemmas, the one most challenging for cooperation is the prisoner's dilemma (PD; see **Glossary**), in which two players choose between cooperating and defecting; cooperation maximizes social welfare, but defection maximizes one's own payoff regardless of the other's choice.

In a well-mixed population in which each individual is equally likely to interact and compete with every other individual, natural selection favors defection in the PD: why should you reduce your own fitness to increase that of a competitor in the struggle for survival? Defectors always out-earn cooperators, and in a population that contains both cooperators and defectors, the latter have higher fitness. Selection therefore reduces the abundance of cooperators until the population consists entirely of defectors. For cooperation to arise, a mechanism for the evolution of cooperation is needed. Such a mechanism is an interaction structure that can cause cooperation to be favored over

defection [1]. These interaction structures specify how the individuals of a population interact to receive payoffs, and how they compete for reproduction. Previous work has identified five such mechanisms for the evolution of cooperation (**Figure 1**): direct reciprocity, indirect reciprocity, spatial selection, multilevel selection, and kin selection. It is important to distinguish between interaction patterns that are mechanisms for the evolution of cooperation and behaviors that require an evolutionary explanation (such as strong reciprocity, upstream reciprocity, and parochial altruism; **Box 2**).

In this article, we build a bridge between theoretical work that has proposed these mechanisms and experimental work exploring how and when people actually cooperate. First we present evidence from experiments that implement each mechanism in the laboratory. Next we discuss why cooperation arises in some experimental settings in which no mechanisms are apparent. Finally, we consider the cognitive underpinnings of human cooperation. We show

Glossary

Evolutionary dynamics: mathematical formalization of the process of evolution whereby a population changes over time. Natural selection operates such that genotypes (or strategies) with higher fitness tend to become more common, whereas lower-fitness genotypes tend to die out. Mutation (re)introduces variation into the population. This process can also represent cultural evolution and social learning, in which people imitate those with higher payoffs and sometimes experiment with novel strategies.

Evolutionary game theory: combination of game theory and evolutionary dynamics. There is a population of agents, each of whom has a strategy. These agents interact with each other and earn payoffs. Payoff is translated into fitness, and the frequency of strategies in the population changes over time accordingly: higher-payoff strategies tend to become more common, whereas lower-payoff strategies tend to die out.

Game theory: mathematical formalization of social interaction and strategic behavior. A given interaction is represented by (i) a set of players, (ii) the choices available to each player, and (iii) the payoff earned by each player depending on both her choice and the choices of the other players. The prisoner's dilemma is one such game that describes the problem of cooperation.

Mechanism for the evolution of cooperation: interaction structure that can cause natural selection to favor cooperation over defection. The mechanism specifies how the individuals of a population interact to receive payoffs, and how they compete for reproduction.

Prisoner's dilemma: game involving two players, each of whom chooses between cooperation or defection. If both players cooperate, they earn more than if both defect. However, the highest payoff is earned by a defector whose partner cooperates, whereas the lowest payoff is earned by a cooperator whose partner defects. It is individually optimal to defect (regardless of the partner's choice) but socially optimal to cooperate. **Box 1** provides further details.

Public goods game: prisoner's dilemma with more than two players. In the public goods game, each player chooses how much money to keep for herself and how much to contribute to an account that benefits all group members.

Corresponding author: Nowak, M.A. (martin_nowak@harvard.edu).

1364-6613/\$ – see front matter

© 2013 Elsevier Ltd. All rights reserved. <http://dx.doi.org/10.1016/j.tics.2013.06.003>



Box 1. Defining cooperation

Consider a game between two strategies, C and D, and the following payoff matrix (indicating the row player's payoff):

	C	D
C	R	S
D	T	P

When does it make sense to call strategy C cooperation and strategy D defection? The following definition [163,164] is useful. The game is a cooperative dilemma if (i) two cooperators obtain a higher payoff than two defectors, $R > P$ yet (ii) there is an incentive to defect. This incentive can arise in three different ways: (a) if $T > R$ then it is better to defect when playing against a cooperator; (b) if $P > S$ then it is better to defect when playing against a defector; and (c) if $T > S$ then it is better to be the defector in an encounter between a cooperator and a defector. If at least one of these three conditions holds, then we have a cooperative dilemma. If none holds, then there is no dilemma and C is simply better than D. If all three conditions hold, we have a prisoner's dilemma, $T > R > P > S$ [6,48].

The prisoner's dilemma is the most stringent cooperative dilemma. Here defectors dominate over cooperators. In a well-mixed population, natural selection always favors defectors over cooperators. For cooperation to arise in the prisoner's dilemma, we need a mechanism for the evolution of cooperation. Cooperative dilemmas that are not the prisoner's dilemma could be called relaxed

cooperative dilemmas. In these games it is possible to evolve some level of cooperation even if no mechanism is at work. One such example is the snowdrift game, given by $T > R > S > P$. Here we find a stable equilibrium between cooperators and defectors, even in a well-mixed population.

If $2R > T + S$, then the total payoff for the population is maximized if everyone cooperates; otherwise a mixed population achieves the highest total payoff. This is possible even for the prisoner's dilemma.

The above definition can be generalized to more than two people (n -person games). We denote by P_i and Q_i the payoffs for cooperators and defectors, respectively, in groups that contain i cooperators and $n-i$ defectors. For the game to be a cooperative dilemma, we require that (i) an all-cooperator group obtains a higher payoff than an all-defector group, $P_n > Q_0$, yet (ii) there is some incentive to defect. The incentive to defect can take the following form: (a) $P_i < Q_{i-1}$ for $i = 1, \dots, n$ and (b) $P_i < Q_i$ for $i = 1, \dots, n-1$. Condition (a) means that an individual can increase his payoff by switching from cooperation to defection. Condition (b) means that in any mixed group, defectors have a higher payoff than cooperators. If only some of these incentives hold, then we have a relaxed cooperative dilemma. In this case some evolution of cooperation is possible even without a specific mechanism. However, a mechanism would typically enhance the evolution of cooperation by increasing the equilibrium abundance of cooperators, increasing the fixation probability of cooperators or reducing the invasion barrier that needs to be overcome. The volunteer's dilemma is an example of a relaxed situation [165]. If all incentives hold, we have the n -person equivalent of a prisoner's dilemma, called the public goods game (PGG) [63], and a mechanism for evolution of cooperation is needed.

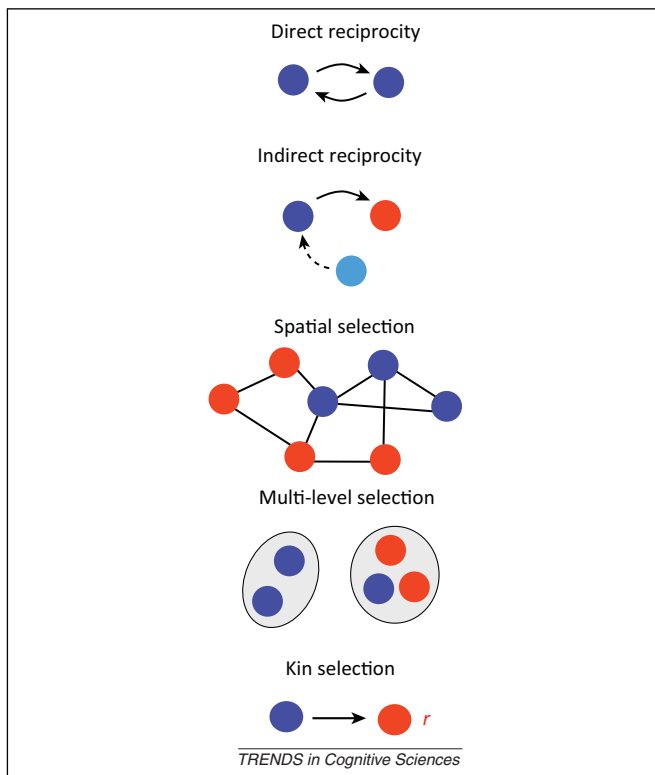


Figure 1. The five mechanisms for the evolution of cooperation. Direct reciprocity operates when two individuals interact repeatedly: it pays to cooperate today to earn your partner's cooperation in the future. Indirect reciprocity involves reputation, whereby my actions towards you also depend on your previous behavior towards others. Spatial selection entails local interaction and competition, leading to clusters of cooperators. Multilevel selection occurs when competition exists between groups and between individuals. Kin selection arises when there is conditional behavior according to kin recognition.

that intuitive, automatic processes implement cooperative strategies that reciprocate, and that these intuitions are affected by prior experience. We argue that these results support a key role for direct and indirect reciprocity in human cooperation, and emphasize the importance of culture and learning.

Five mechanisms

Direct reciprocity

Direct reciprocity arises if there are repeated encounters between the same two individuals [2–5]. Because they interact repeatedly, these individuals can use conditional strategies whereby behavior depends on previous outcomes. Direct reciprocity allows the evolution of cooperation if the probability of another interaction is sufficiently high [6]. Under this 'shadow of the future', I may pay the cost of cooperation today to earn your reciprocal cooperation tomorrow. The repeated game can occur with players making simultaneous decisions in each round or taking turns [7]. Successful strategies for the simultaneous repeated PD include tit-for-tat (TFT), a strategy that copies the opponent's previous move, and win-stay lose-shift, a strategy that switches its action after experiencing exploitation or mutual defection [8]. TFT is an excellent catalyst for the emergence of cooperation, but when errors are possible it is quickly replaced by strategies that sometimes cooperate even when the opponent defects (e.g., Generous TFT) [9].

Indirect reciprocity

Indirect reciprocity operates if there are repeated encounters within a population and third parties observe some of these encounters or find out about them. Information about

Box 2. Behavioral patterns versus mechanisms for the evolution of cooperation

It is important to distinguish mechanisms for the evolution of cooperation from behavioral patterns that are not themselves mechanisms. Three examples are upstream reciprocity, strong reciprocity, and parochial altruism. Upstream (or generalized) reciprocity refers to the phenomenon of paying it forward, by which an individual who has just received help is more likely to help others in turn. Strong reciprocity refers to individuals who reward cooperation and punish selfishness, even in anonymous interactions with no promise of future benefits. Parochial altruism (or in-group bias) describes the behavior whereby people are more likely to help members of their own group than members of other groups.

None of these concepts explains the evolution of cooperation: adding one or more of these elements to a prisoner's dilemma will not cause selection to favor cooperation. Instead, these concepts are descriptions of behavior that require an evolutionary explanation. Group selection, spatial structure, or some chance of direct or indirect reciprocity can lead to the evolution of upstream reciprocity [166,167], strong reciprocity [13,39,168], and parochial altruism [122,139,169–171].

such encounters can spread through communication, affecting the reputations of the participants. Individuals can thus adopt conditional strategies that base their decision on the reputation of the recipient [10,11]. My behavior towards you depends on what you have done to me and to others. Cooperation is costly but leads to the reputation of being a helpful individual, and therefore may increase your chances of receiving help from others. A strategy for indirect reciprocity consists of a social norm and an action rule [12–14]. The social norm specifies how reputations are updated according to interactions between individuals. The action rule specifies whether or not to cooperate given the available information about the other individual. Indirect reciprocity enables the evolution of cooperation if the probability of knowing someone's reputation is sufficiently high.

Spatial selection

Spatial selection can favor cooperation without the need for strategic complexity [15,16]. When populations are structured rather than randomly mixed, behaviors need not be conditional on previous outcomes. Because individuals interact with those near them, cooperators can form clusters that prevail, even if surrounded by defectors. The fundamental idea is that clustering creates assortment whereby cooperators are more likely to interact with other cooperators. Therefore, cooperators can earn higher payoffs than defectors. More generally, population structure affects the outcome of the evolutionary process, and some population structures can lead to the evolution of cooperation [17,18]. Population structure specifies who interacts with whom to earn payoffs and who competes with whom for reproduction. The latter can be genetic or cultural. Population structure can represent geographic distribution [19,20] or social networks [21], and can be static [22–24] or dynamic [21,25–29]. Population structure can also be implemented through tag-based cooperation, in which interaction and cooperation are determined by arbitrary tags or markers [30–32]. In this case, clustering is not literally spatial but instead occurs in the space of phenotypes [30].

Multilevel selection

Multilevel selection operates if, in addition to competition between individuals in a group, there is also competition between groups [33–39]. It is possible that defectors win within groups, but that groups of cooperators outcompete groups of defectors. Overall, such a process can result in the selection of cooperators. Darwin wrote in 1871: 'There can be no doubt that a tribe including many members who . . . were always ready to give aid to each other and to sacrifice themselves for the common good, would be victorious over other tribes; and this would be natural selection.' [40].

Kin selection

Kin selection can be seen as a mechanism for the evolution of cooperation if properly formulated. In our opinion, kin selection operates if there is conditional behavior based on kin recognition: an individual recognizes kin and behaves accordingly. As J.B.S. Haldane reportedly said, 'I will jump into the river to save two brothers or eight cousins' [41]. Much of the current literature on kin selection, however, does not adhere to this simple definition based on kin recognition. Instead, kin selection is linked to the concept of inclusive fitness [42]. Inclusive fitness is a particular mathematical method to account for fitness effects. It assumes that personal fitness can be written as a sum of additive components caused by individual actions. Inclusive fitness works in special cases, but makes strong assumptions that prevent it from being a general concept [43]. A straightforward mathematical formulation describing the evolutionary dynamics of strategies or alleles without the detour of inclusive fitness is a more universal and more meaningful approach. This position critical of inclusive fitness, which is based on a careful mathematical analysis of evolution [43], has been challenged by proponents of inclusive fitness [44], but without considering the underlying mathematical results [45]. In our opinion, a clear understanding of kin selection can only emerge once the intrinsic limitations of inclusive fitness are widely recognized. Meanwhile, it is useful to remember that no phenomenon in evolutionary biology requires an inclusive fitness-based analysis [43].

Interactions between mechanisms

Each of these mechanisms applies to human cooperation. Over the course of human evolution, it is likely that they were (and are) all in effect to varying degrees. Although each mechanism has traditionally been studied in isolation, it is important to consider the interplay between them. In particular, when discussing the evolution of any prosocial behavior in humans, we cannot exclude direct and indirect reciprocity. Early human societies were small, and reputation and reputation were always in play. Even in the modern world, most of our crucial interactions are repeated, such as those with our coworkers, friends, and family. Thus, spatial structure, group selection, and kin selection should be considered in the context of their interactions with direct and indirect reciprocity. Surprising dynamics can arise when mechanisms are combined. For example, direct reciprocity and spatial structure can interact either synergistically or antagonistically, depending on the levels of repetition and assortment [46]. Further

exploration of the interactions between mechanisms is a promising direction for future research.

Experimental evidence in support of the five mechanisms

Theoretical work provides deep insights into the evolution of human cooperation. Evolutionary game theory allows us to explore what evolutionary trajectories are possible and what conditions may give rise to cooperation. To investigate how cooperation among humans in particular arises and is maintained, theory must be complemented with empirical data from experiments [47]. Theory suggests what to measure and how to interpret it. Experiments illuminate human cooperation in two different ways: by examining what happens when particular interaction structures are imposed on human subjects, and by revealing the human psychology shaped by mechanisms that operate outside of the laboratory (Box 3).

We now present both types of experimental evidence. First we describe experiments designed to test each of the mechanisms for the evolution of cooperation in the laboratory. We then discuss the insights gained from cooperation in one-shot anonymous experiments. For comparability with theory, we focus on experiments that study cooperation using game theoretic frameworks. Most of these experiments are incentivized: the payout people receive depends on their earnings in the game. Subjects are told the true rules of the game and deception is prohibited: to explore the effect of different rules on cooperation, subjects must believe that the rules really apply. Finally, interactions are typically anonymous, often occurring via computer terminals or over the internet. This anonymity reduces concerns about reputational effects outside of the laboratory, creating a baseline from which to measure the effect of adding more complicated interaction structures.

Box 3. How behavioral experiments inform evolutionary models

Experiments shed light on human cooperation in different ways [47]. One type of experiment seeks to recreate the rules of interaction prescribed by a given model. By allowing human subjects to play the game accordingly, researchers test the effect of adding human psychology. Do human agents respond to the interaction rules similarly to the agents in the models? Or are important elements of proximate human psychology missing from the models, revealing new questions for evolutionary game theorists to answer?

Other studies explore behavior in experiments in which no mechanisms that promote cooperation are present (e.g., one-shot anonymous games in well-mixed populations). By examining play in these artificial settings, we hope to expose elements of human psychology and cognition that would ordinarily be unobservable. For example, in repeated games, it can be self-interested to cooperate. When we observe people who cooperate in repeated games, we cannot tell if they have a predisposition towards cooperation or are just rational selfish maximizers. One-shot anonymous games are required to reveal social preferences. The artificiality of these laboratory experiments is therefore not a flaw, but can make such experiments valuable. It is critical, however, to bear this artificiality in mind when interpreting the results: these experiments are useful because of what they reveal about the psychology produced by the outside world, rather than themselves being a good representation of that world.

Direct reciprocity

Over half a century of experiments [48] demonstrate the power of repetition in promoting cooperation. Across many experiments using repeated PDs, people usually learn to cooperate more when the probability of future interaction is higher [49–55] (in these games, there is typically a constant probability that a given pair of subjects will play another round of PD together). Repetition continues to support cooperation even if errors are added (the computer sometimes switches a player's move to the opposite of what she intended) [55], which is consistent with theoretical results [9,56]. More quantitatively, theoretical work using stochastic evolutionary game theory (modeling that incorporates randomness and chance) finds that cooperation will be favored by selection if TFT earns a higher payoff than the strategy Always Defect (ALLD) in a population in which the two strategies are equally common (when TFT is risk-dominant over ALLD) [57]. More generally, as the payoff for TFT relative to ALLD in such a mixed population increases, so too does the predicted frequency of cooperation. Here we show that this prediction does an excellent job of organizing the experimental data: across 14 conditions from four papers, the fraction of cooperators is predicted with $R^2 = 0.81$ by the extent to which the probability of future interaction exceeds the risk dominance threshold (Figure 2). This is one of numerous situations in which stochastic evolutionary game theory [57] successfully describes observed human behavior [58–61].

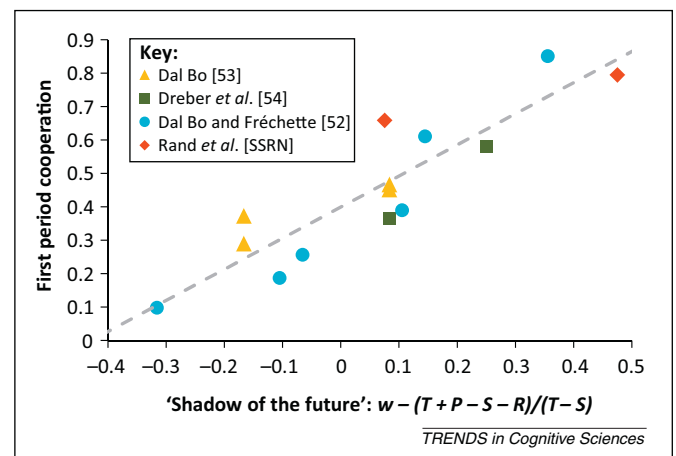


Figure 2. Repetition promotes cooperation in the laboratory. The frequency of cooperative strategies in various repeated prisoner's dilemma (PD) experiments is plotted as a function of the extent to which future consequences exist for actions in the current period. Specifically, the x-axis shows the amount by which the continuation probability w (probability that two subjects play another PD round together) exceeds the critical payoff threshold $(T + P - S - R)/(T - S)$ necessary for tit-for-tat (TFT) to risk-dominate always defect (ALLD). In a population that is 1/2 TFT and 1/2 ALLD, $w < (T + P - S - R)/(T - S)$ means that ALLD earns more than TFT; $w = (T + P - S - R)/(T - S)$ means that TFT and ALLD do equally well; and the more w exceeds $(T + P - S - R)/(T - S)$, the more TFT earns compared to ALLD. The y-axis indicates the probability of cooperation in the first round of each repeated PD game (cooperation in the first period is a pure reflection of one's own strategy, whereas play in later periods is influenced by the partner's strategy as well). Data are from [52–54] and [Rand, D.G., et al. (2013) It's the thought that counts: the role of intentions in reciprocal altruism, <http://ssrn.com/abstract=2259407>]. For maximal comparability, we do not include the treatments from [54] with costly punishment, or the treatments from Rand et al. (<http://ssrn.com/abstract=2259407>) with exogenously imposed errors. Owing to variations in experimental design, subjects in different experiments had differing lengths of time to learn. Nonetheless, a clear increasing relationship is evident, both within each study and over all studies. The trend line shown is given by $y = 0.93x + 0.40$, with $R^2 = 0.81$.

Repetition promotes cooperation in dyadic interactions. The situation is more complicated, however, if groups of players interact repeatedly [62]. Such group cooperation is studied in the context of the public goods game (PGG) [63], an n -player PD. The PGG is typically implemented by giving each of n players an endowment and having them choose how much to keep for themselves and how much to contribute to the group. All contributions are multiplied by some constant r ($1 < r < n$) and split equally by all group members. The key difference from the two-player PD is that in the PGG, targeted interactions are not possible: if one player contributes a large amount while another contributes little, a third group member cannot selectively reward the former and punish the latter. The third player can choose either a high contribution, rewarding both players, or a low contribution, punishing both. Thus, although direct reciprocity can in theory stabilize cooperation in multiplayer games, this stability is fragile and can be undermined by errors or a small fraction of defectors [64]. As a result, cooperation almost always fails in repeated PGGs in the laboratory [65–67].

Does this mean that mechanisms other than direct reciprocity are needed to explain group cooperation? The answer is no. We must only realize that group interactions do not occur in a vacuum, but rather are superimposed on a network of dyadic personal relationships. These personal, pairwise relationships allow for the targeted reciprocity that is missing in the PGG, giving us the power to enforce group-level cooperation. They can be represented by adding pairwise reward or punishment opportunities to the PGG. (Box 4 discusses costly punishment in repeated two-player games). After each PGG round, subjects can pay to increase or decrease the payoff of other group members according to their contributions. Thus, the possibility of targeted interaction is reintroduced, and direct reciprocity can once again function to promote cooperation.

Numerous laboratory experiments demonstrate that pairwise reward and punishment are both effective in promoting cooperation in the repeated PGG [65–70]. Naturally, given that both implementations of direct

reciprocity promote cooperation, higher payoffs are achieved when using reward (which creates benefit) than punishment (which destroys it). Rewarding also avoids vendettas [54,71] and the possibility of antisocial punishment, whereby low contributors pay to punish high contributors. It has been demonstrated that antisocial punishment occurs in cross-cultural laboratory experiments [72–74] and can prevent the evolution of cooperation in theoretical models [75–78]. These cross-cultural experiments add a note of caution to previous studies on punishment and reward in the PGG: targeted interactions can only support cooperation if they are used properly. Antisocial punishment undermines cooperation, as does rewarding of low contributors [Ellingsen, T. *et al.* (2012) Civic capital in two cultures: the nature of cooperation in Romania and USA, <http://ssrn.com/abstract=2179575>]. With repetition and the addition of pairwise interactions, cooperation can be a robust equilibrium in the PGG, but populations can nonetheless become stuck in other, less efficient equilibria or fail to equilibrate at all.

Taken together, the many experiments exploring the linking of dyadic and multiplayer repeated games demonstrate the power of direct reciprocity for promoting large-scale cooperation. Interestingly, this linking also involves indirect reciprocity: if I punish a low contributor, then I reciprocate a harm done to me (direct reciprocity) as well as a harm done to other group members (indirect reciprocity [79]). Further development of theoretical models analyzing linked games is an important direction for future research, as is exploring the interplay between direct and indirect reciprocity in such settings.

Indirect reciprocity

Indirect reciprocity is a powerful mechanism for promoting cooperation among subjects who are not necessarily engaged in pairwise repeated interactions. To study indirect reciprocity in the laboratory, subjects typically play with randomly matched partners and are informed about their choices in previous interactions with others [80,81]. Most subjects condition their behavior on this information: those who have been cooperative previously, particularly towards partners who have behaved well themselves, tend to receive more cooperation [80–89]. Thus, having a reputation of being a cooperator is valuable, and cooperation is maintained: it is worth paying the cost of cooperation today to earn the benefits of a good reputation tomorrow. Figure 3 provides quantitative evidence of the value subjects place on a good reputation by linking PD games with a market in which reputation can be bought and sold [82].

It has also been shown that reputation effects promote prosocial behavior outside of the laboratory. Field experiments find that publicizing the names of donors increases the level of blood donation [90] and giving to charity [91]. It was also shown that non-financial incentives involving reputation outperformed monetary incentives in motivating participation in an energy blackout prevention program in California [92] and the sale of condoms on behalf of a health organization in Namibia [Ashraf, N. *et al.* (2012) No margin, no mission? A field experiment on incentives for pro-social tasks, Harvard Business School Working Paper].

Box 4. Tit-for-tat versus costly punishment

The essence of direct reciprocity is that future consequences exist for present behavior: if you do not cooperate with me today, I will not cooperate with you tomorrow. This form of punishment, practiced by TFT in pairwise interactions, via denial of future reward is different from costly punishment; in the latter case, rather than just defecting against you tomorrow, I actually pay a cost to impose a cost on you [54,65–67,84,172–175].

The following question therefore arises: what is the role of costly punishment in the context of repeated pairwise interactions? A set of behavioral experiments revealed that costly punishing in the repeated PD was disadvantageous, with punishers earning lower payoffs than non-punishers. This was because punishment led to retaliation much more often than to reconciliation [54]. Complementing these observations are evolutionary simulations that revealed similar results: across a wide range of parameter values, selection disfavors the use of costly punishment in the repeated PD [61]. Similar results were found in an evolutionary model based on group selection [176]: even a minimal amount of repetition in which a second punishment stage is added causes selection to disfavor both punishment and cooperation because of retaliation.

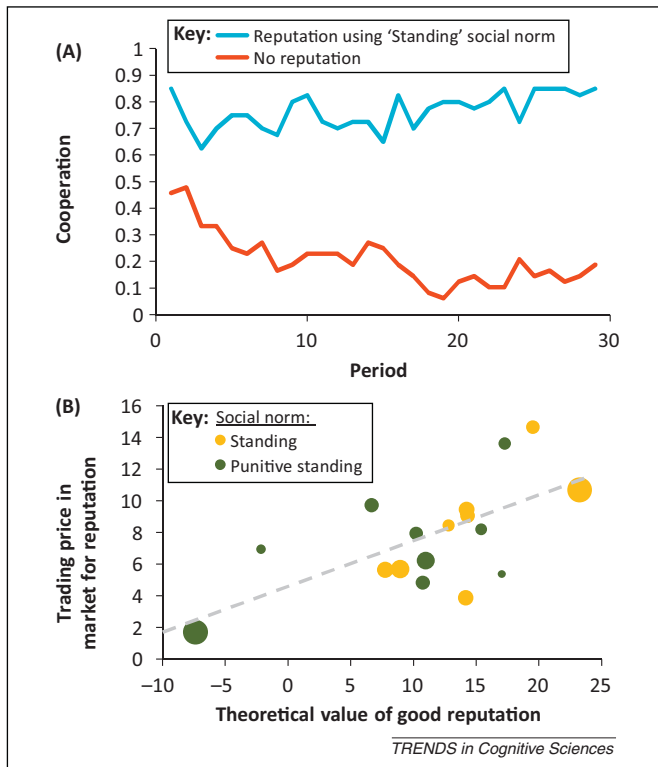


Figure 3. Formal reputation systems make cooperation profitable. **(A)** In a series of randomly shuffled PDs without reputation, cooperation decays over time. In the reputation condition, however, cooperation is maintained at a high rate. Here, subjects are assigned a label of 'good' or 'bad' in each round, depending on their behavior. The social norm referred to as 'standing' is used: Cooperating gives a good reputation and defecting gives a bad reputation, except when a good player meets a bad player; in this case, the good player must defect to obtain a good reputation. **(B)** Cooperation is costly, but you can benefit from the good reputation you receive if it increases the chance that others will cooperate with you in the future. Thus, the more people in a particular group are inclined to cooperate with those with a good reputation, the greater the value of having a good reputation in that group. Allowing people to buy and sell reputations in a market can be used to assess whether people explicitly understand the value of a good reputation. As is shown here, there is a strong positive correlation between the theoretical value of a good reputation in a given group and the equilibrium trading price in the market (each circle represents one group, with size proportional to the total number of trades in the market). This positive relationship exists using both standing and an alternate norm in which two players with a bad reputation must defect with each other to regain a good reputation. Data reproduced from [82].

Indirect reciprocity relies on peoples' ability to effectively communicate and distribute reputational information. Not surprisingly, people spend a great deal of their time talking to each other (gossiping) about the behavior of third parties [85,93]. In addition to this traditional form of transmitting reputational information, the internet has dramatically expanded our ability to maintain large-scale reputation systems among strangers. For example, online markets such as eBay have formalized reputation systems in which buyers rate sellers. As predicted by indirect reciprocity, there is a large economic value associated with having a good eBay reputation [94]. Similarly, business rating websites such as Yelp.com create a global-level reputation system, allowing people without local information to reliably avoid low-quality products and services, and creating economic incentives for businesses to earn good reputations [Luca, M. (2011) Reviews, reputation, and revenue: the case of Yelp.com, Harvard Business School NOM Unit Working Paper].

A fascinating question that these studies raise is why people bother to leave evaluations at all. Or, even when people do provide information, why be truthful? Providing accurate information requires time and effort, and is vital for reputation systems to function. Thus, rating is itself a public good [95]. However, indirect reciprocity may be able to solve this second-order free-rider problem itself: to remain in good reputation, you must not only cooperate in the primary interactions but also share truthful information. Exploring this possibility further is an important direction for future research.

Enforcement poses another challenge for indirect reciprocity. Withholding cooperation from defectors is essential for the reputation system to function. However, doing so can potentially be damaging for your own reputation. This is particularly true when using simple reputation systems such as image scoring [10], which is a first-order assessment rule that only evaluates actions (cooperation is good, defection is bad). However, it can apply even when using more complex reputation rules whereby defecting against someone with a bad reputation earns you a good reputation: if observers are confused about the reputation of your partner, defecting will tarnish your name. Here we suggest a possible solution to this problem. If players have the option to avoid interacting with others, they may shun those in bad reputation. Thus, they avoid being exploited while not having to defect themselves. Such a system should lead to stable cooperation using even the simplest of reputation systems. Another interesting possibility involves intermediation: if you employ an intermediary to defect against bad players on your behalf, this may help to avoid sully your reputation. Consistent with this possibility, experimental evidence suggests that the use of intermediaries reduces blame for selfish actions [96,97]. We expect that researchers will explore these phenomena further in the coming years, using theoretical models as well as laboratory and field experiments.

Finally, there is evidence of the central role of reputational concerns in human evolution. Infants as young as 6 months of age take into account others' actions toward third parties when making social evaluations [98,99]. This tendency even occurs between species: capuchin monkeys are less likely to accept food from humans who were unhelpful to third parties [100]. Humans are also exquisitely sensitive to the possibility of being observed by third parties [101]. For example, people are more prosocial when being watched by a robot with large fake eyes [102] or when a pair of stylized eye-spots is added to the desktop background of a computer [103]. In the opposite direction, making studies double-blind such that experimenters cannot associate subjects with their actions increases selfishness [104].

Spatial selection

Unlike direct and indirect reciprocity, experimental evidence in support of spatial selection among humans is mixed. (There is good evidence for spatial selection in unicellular organisms [105]). Experiments that investigate fixed spatial structures typically assign subjects to locations in a network and have them play repeatedly with their neighbors. Cooperation rates are then compared to a

control in which subjects' positions in the network are randomly reshuffled in each round, creating a well-mixed population. As in theoretical models, subjects in these experiments are usually given a binary choice, either cooperate with all neighbors or defect with all neighbors; and are typically presented in each round with the payoff and choice of each neighbor. However, unlike the models, cooperation rates in these experiments are no higher in structured than in well-mixed populations [106–110].

Various explanations have been advanced for this surprising set of findings. One suggestion is that subjects in laboratory experiments engage in high rates of experimentation, often changing their strategies at random rather than copying higher-payoff neighbors [108]. Such experimentation is analogous to mutation in evolutionary models. High mutation rates undermine the effect of spatial structure: when players are likely to change their strategies at random, then the clustering that is essential for spatial selection is disrupted [111]. Without sufficient clustering, cooperation is no longer advantageous.

Another explanation involves the way in which subjects choose which strategy to adopt. Theoretical models make detailed assumptions about how individuals update their strategies, and whether network structure can promote cooperation depends critically on these details [18]. It is possible that human subjects in the experimental situations examined thus far tend to use update rules that cancel the effect of spatial structure [108]. A related argument involves the confounding of spatial structure and direct reciprocity that occurs in these experiments [112]. Subjects in the experiments know that they are interacting repeatedly with the same neighbors. Thus, they can play conditional strategies, unlike the agents in most theoretical models. Because players must choose the same action towards all neighbors, players in these experiments cannot target their reciprocity (like in the PGG). Thus, a tendency to reciprocate may lead to the demise of cooperation.

Here we offer a possible alternative explanation. Theoretical work has provided a simple rule for when a fixed network structure will promote cooperation: cooperation is only predicted to be favored when the PD benefit-to-cost ratio exceeds the average number of neighbors in the network [23]. In most of the experiments on fixed networks to date, this condition is not satisfied. Thus, it remains possible that fixed networks will actually succeed in promoting cooperation for the right combinations of payoffs and structure. Exploring this possibility is an important direction for future study.

In contrast to these negative results using static networks, dynamic networks successfully promote cooperation in the laboratory (Figure 4) [113–116]. In these experiments, subjects can make or break connections with others and the network evolves over time. This dynamic nature allows subjects to engage in targeted action via 'link reciprocity': players can choose to sever links with defectors or make links with cooperators. The importance of dynamic assortment based on arbitrary tags has also been demonstrated in laboratory experiments using coordination games: associations between tags and actions emerge spontaneously, as does preferential interaction between players sharing the same tag [117].

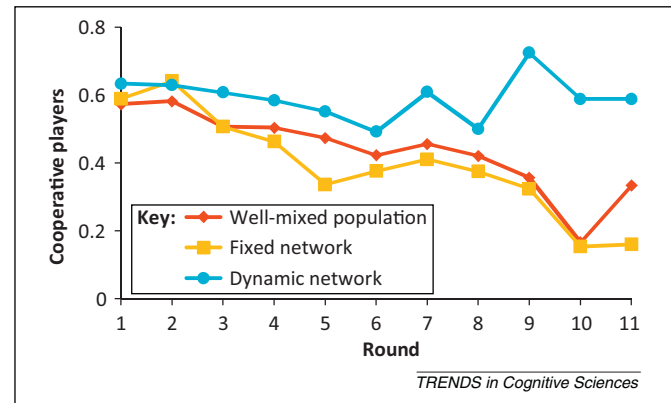


Figure 4. In behavioral experiments, dynamic social networks can promote cooperation via link reciprocity. The fraction of subjects cooperating in a multilateral cooperation game is shown (cooperation entailed paying 50 units per neighbor for all neighbors to gain 100 units). In the well-mixed condition, the network was randomly shuffled in every round. In the fixed network condition, subjects interacted with the same neighbors in each round. In the dynamic network condition, 30% of player pairs were selected at random, and one of the two players could unilaterally update the connection (i.e., break an existing link or create a link if none existed before). Data reproduced from [113].

More generally, there is substantial evidence that social linkages and identity are highly flexible. Minimal cues of shared identity (such as preference for similar types of paintings, i.e., the minimal groups paradigm) can increase cooperation among strangers [118]. Alternatively, introduction of a higher-level threat can realign coalitions, making yesterday's enemies into today's allies [119,120]. Such plasticity is not limited to modern humans: many early human societies were characterized by fission–fusion dynamics, whereby group membership changed regularly [121]. The development of evolutionary models that capture this multifaceted and highly dynamic nature of group identity is a promising direction for future work. Models based on changing set memberships [27,122] and tag-based cooperation [30–32] represent steps in this direction.

Finally, studies examining behavior in real-world networks also provide evidence of the importance of population structure in cooperation. For example, experiments with hunter–gatherers show that social ties predict similarity in cooperative behavior [123]. A nationally representative survey of American adults found that people who engage in more prosocial behavior have more social contacts, as predicted by dynamic network models [124]. There is also evidence that social structure is heritable [125], as is assumed in many network models.

In sum, there is evidence that spatial selection is an important force in at least some domains of human cooperation. However, further work is needed to clarify precisely when and in which ways spatial selection promotes cooperation in human interactions.

Multilevel selection

In the laboratory, multilevel selection is typically implemented using interaction structures in which groups compete with each other. For example, two groups play a PGG and compete over a monetary prize: the group with the larger total contribution amount wins, and each member of that group shares equally in the prize. Thus, the incentive to defect in the baseline PGG is reduced by the potential gain from winning the group competition, although

Box 5. In-group bias is not necessarily evidence of selection at the level of the group

Some might argue that the ubiquitousness of in-group bias is proof that multilevel selection played a central role in human evolution. In-group bias, or parochial altruism, is a behavioral pattern whereby people cooperate more with members of their own group than with out-group members [118,119,177,178]. It is true that multilevel selection and inter-group conflict can lead to in-group bias [139,169]. However, other mechanisms can also give rise to in-group bias. Spatial selection can lead to the evolution of in-group bias via set-structured interactions or tag-based cooperation [30,121,171]. Reciprocity can also favor in-group bias. For example, in the context of direct reciprocity, it seems likely that the probability of future interaction is greater for in-group than for out-group members. Given this, it could be adaptive to play cooperative strategies such as TFT with in-group members but to play ALLD with out-group members. Similarly, in the context of indirect reciprocity, information about the behavior of out-group members may be less accurate or detailed [170]. Thus, the presence of in-group bias in human psychology can be explained by different mechanisms and does not necessarily indicate multilevel selection.

defection is typically still the payoff-maximizing choice. Numerous such experiments have shown that competition between groups increases cooperation substantially [126–131]. Furthermore, just phrasing the interaction as a competition between groups, without any monetary prize for winning, also increases cooperation [130,132]. Experience with real-world intergroup conflict also increases cooperation [133,134]. (Note that although the prevalence of in-group favoritism may seem to indicate a psychology shaped by intergroup conflict, such bias can also be explained by other mechanisms; Box 5). In sum, there is ample evidence that intergroup competition can be a powerful force for promoting within-group cooperation.

Critics of multilevel selection argue that empirically, the conditions necessary for substantial selection pressure at the group level were not met over the course of human history [135]: concerns include low ratios of between-group to within-group variation because of factors such as migration and mutation/experimentation, and the infrequency of group extinction or lethal inter-group warfare. The laboratory experiments discussed above do not address these concerns: in these studies, the interaction structure is explicitly constructed to generate group-level selection. Instead, anthropological and archaeological data have been used to explore when the conditions necessary for multilevel selection have been satisfied in human history, either at the genetic [37,38] or cultural [136] level.

Kin selection

Perhaps surprisingly, kin selection is the least-studied mechanism for human cooperation. Research on humans largely focuses on cooperation between non-kin. In part this is because cooperation between related individuals is seen as expected and therefore uninteresting. Furthermore, humans cooperate with unrelated partners at a much higher rate than for other species, and thus non-kin cooperation is an element of potential human uniqueness. There are also substantial practical hurdles to studying kin selection in humans. The effect of kinship is difficult to measure, because relatedness and reciprocity are inexorably intertwined: we

almost always have long-lasting reciprocal relationships with our close genetic relatives.

Nonetheless, understanding the role of kinship in the context of human cooperation is important. Parents helping children is not an example of kin selection, but rather straightforward selection-maximizing direct fitness. Kin selection, however, may be at work in interactions between collateral kin (family members who are not direct descendants). In this context, some scholars have investigated the cues used for kin recognition. For example, in predicting self-reported altruistic behavior, an interaction has been found between observing your mother caring for a sibling (maternal perinatal association, MPA) and the amount of time spent living with a sibling (co-residence) [137]: MPA is a strong signal of relatedness, and thus co-residence does not predict altruism in the presence of MPA. In the absence of MPA (e.g., if you are a younger sibling who did not observe your older siblings being cared for), however, co-residence does predict altruism. This interaction suggests that co-residence is used as an indication of relatedness, rather than only as an indication of the probability of future interaction.

More studies on this topic are needed, in particular the development of experiments that tease apart the roles of kinship and reciprocity. Progress in this area would be aided by theoretical developments combining evolutionary game theory and population genetics [43].

Cooperation in the absence of any mechanisms

How can we explain cooperation in one-shot anonymous laboratory games between strangers? Such cooperation is common [138], yet seems to contradict theoretical predictions because none of the five mechanisms appears to be in play: no repetition or reputation effects exist, interactions are not structured, groups are not competing, and subjects are not genetic relatives. Yet many subjects still cooperate. Why? Because the intuitions and norms that guide these decisions were shaped outside the laboratory by mechanisms for the evolution of cooperation.

How exactly this happens is a topic of debate. There are two dimensions along which scholars disagree: (i) whether cooperation in one-shot interactions is explicitly favored by evolution (through spatial or multilevel selection) or is the result of overgeneralizing strategies from settings in which cooperation is in one's long-run self-interest (due to direct and indirect reciprocity); and (ii) the relative importance of genetic evolution versus cultural evolution in shaping human cooperation.

On the first dimension, one perspective argues that multilevel selection and spatial structure specifically favor altruistic preferences that lead to cooperation in one-shot anonymous settings [38,39,139]. Thus, although laboratory experiments may not explicitly include these effects, they have left their mark on the psychology that subjects bring into the laboratory by giving rise to altruism. The alternative perspective argues that direct and indirect reciprocity were the dominant forces in human evolution. By this account, selection favors cooperative strategies because most interactions involve repetition or reputation. Because cooperation is typically advantageous, we internalize it as our default behavior. This

cooperative predisposition is then sometimes overgeneralized, spilling over into unusual situations in which others are not watching [103,140]. In this view, cooperation in anonymous one-shot settings is a side effect of selection for reciprocal cooperation, rather than an active target of selection itself. Note that in both views, evolution gives rise to people who are truly altruistic and cooperate even when there are no future benefits from doing so: the disagreement is over whether or not that altruism was directly favored by selection or is a byproduct of selection in non-anonymous interactions.

Turning to the second dimension, all of the mechanisms for the evolution of cooperation can function via either genetic or cultural evolution. In the context of cultural evolution, traits spread through learning, often modeled as imitation of strategies that yield higher payoffs or are more common [141]. It has been argued by some that multilevel selection promotes cooperation through genetic evolution [36], whereas others posit an important role of culture [38,142–144]. The same is true for reciprocity. We might have genetic predispositions to cooperate because our ancestors lived in small groups with largely repeated interactions [140,145]. Or we might have learned cooperation as a good rule of thumb for social interaction, because most of our important relationships are repeated and thus cooperation is typically advantageous, as per the ‘social heuristics hypothesis’ [146] [Rand, D.G. *et al.* (2013) Intuitive cooperation and the social heuristics hypothesis: evidence from 15 time constraint studies, <http://ssrn.com/abstract=2222683>]. Thus one’s position in this second area of debate need not be tied to one’s belief about the first.

Intuitive reciprocation

To help distinguish between these different possibilities, we examine the cognitive basis of cooperation. Experiments using economic games have shown that automatic, intuitive processes support cooperation in one-shot games, whereas reflection and deliberation lead to selfishness. Faster decisions in the PGG tend to be more cooperative [146] (Figure 5A). Induction of an intuitive mindset through priming or time pressure increases PGG cooperation relative to a more reflective mindset [146,147]. Increasing the role of intuition through cognitive load augments generosity in a resource allocation game [148] and in a unilateral money division task (i.e., dictator game [149,150]). Affective, emotional responses play an important role in prosocial decision-making [151–153]. These findings suggest that cooperation in one-shot anonymous interactions involves some overgeneralization: more generalized processes involving intuition and emotion favor the typically advantageous behavior of cooperation, whereas reflection and reasoning adjust towards the behavior that is payoff-maximizing in the specific context of one-shot games (i.e., selfishness).

These experiments support the argument that cooperative strategies develop in the context of direct and indirect reciprocity, and are then misapplied to one-shot games. We now evaluate a further prediction of this line of reciprocity-based reasoning: cooperation should not always be intuitive. A key element of direct and indirect reciprocity is conditional cooperation. As exemplified by the TFT

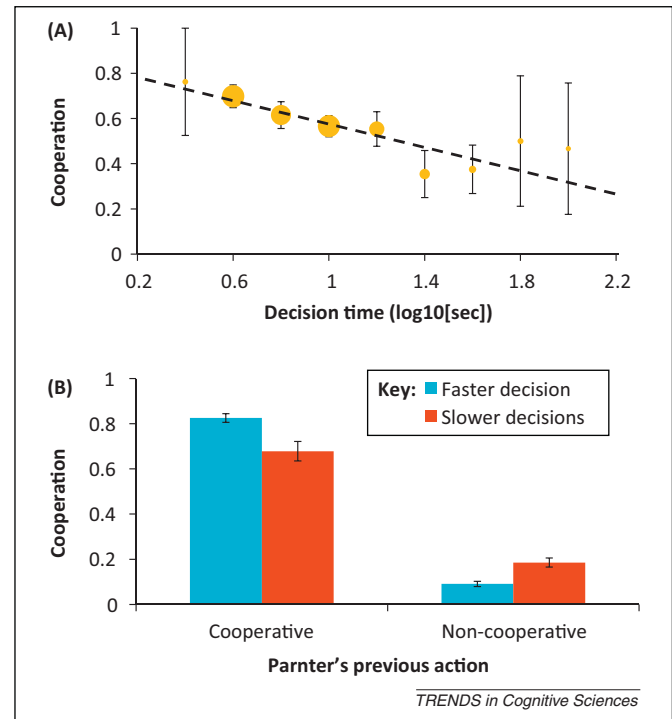


Figure 5. Automatic, intuitive responses involve reciprocal cooperation strategies. (A) In a one-shot public good game, faster decisions are more cooperative. Thus, it is intuitive to cooperate in anonymous settings. Data reproduced from [146]. (B) In a repeated prisoner’s dilemma, faster decisions are more cooperative when the partner cooperated in the previous round, and are less cooperative when the partner did not cooperate in the previous round. Thus, it is intuitive to reciprocate in repeated settings. Analysis of data from [54] and the no-error condition of [55]. For visualization, we categorize decisions made in <2 s as faster and decisions made in ≥ 2 s as slower; however our regression analysis treats decision time as a continuous variable.

strategy, reciprocal interactions should lead to intuitions that favor cooperation at the outset of a relationship, and cooperation in response to a cooperative partner. However in response to a selfish partner, the automatic response should reverse to selfishness. Put differently, reciprocity-based hypotheses for the evolution of human cooperation predict intuitive reciprocation.

To evaluate this prediction, we reanalyze decision time data from two experiments in which subjects played repeated PDs [54,55]. We ask how the relationship between decision time and cooperation varies according to the partner’s previous move (Figure 5B). As predicted, we find that if the partner cooperated in the previous round, faster decisions are significantly more cooperative ($P < 0.001$), but if the partner did not cooperate in the previous round, faster decisions are significantly less cooperative ($P = 0.004$). A regression predicting cooperation across both situations shows a significant interaction between decision time and partner’s previous move ($P < 0.001$). All p values were generated using logistic regression with robust standard errors clustered on subject and pairing, including controls for PD payoff specification (benefit-to-cost ratio), continuation probability, presence of a costly punishment option, game number, and round number; decision times were \log_{10} -transformed, as in [146]; treatments in [55] with exogenously imposed execution errors are not included, because they change the strategic nature of the interaction; in total, 8509 decisions by 152 subjects were analyzed.

Further evidence for the intuitive reciprocation prediction comes from the ultimatum game (UG). In the UG, one player (the proposer) makes an offer of how to split a sum of money with a second player (the responder). If the responder rejects, neither receives anything. Both behavioral experiments and neuroimaging studies suggest that when responders are confronted with unfair offers, the intuitive decision is to reject, whereas reflection leads to increased acceptance [154–157] (although evidence from transcranial magnetic stimulation experiments suggests that deliberative processes also play some role in rejections [158,159]). Thus, intuition again favors reciprocation (in this case, paying a cost to retaliate against selfishness). Like cooperation in one-shot interactions, rejection of unfair offers in the UG is not payoff-maximizing in the one-shot games studied in the laboratory, but is adaptive in the context of reciprocal interactions [160].

This evidence of intuitive reciprocation supports the argument that strategies selected in the context of repeated games spill over into one-shot anonymous interactions. Are these intuitions the result of genetic hard-coding or of learning and experience? Several additional results support the latter hypothesis. Some experiments find no effect of promoting intuition on cooperative behavior in one-shot games [147,161,162], suggesting that cooperative intuitions are not universal. Specific moderators of the intuitive cooperation effect have also been demonstrated. One-shot cooperation is only intuitive among people from communities in which most others are trustworthy and cooperative themselves [146]. If you grow up in a non-cooperative equilibrium in which cooperation is not payoff-maximizing, you internalize defection as your default. Prior experience with behavioral experiments also moderates the role of intuition in cooperation. An individual-differences study showed that intuitive responses are more cooperative among naïve subjects, but that intuition does not promote cooperation among experienced subjects [146]. At the study level, the effect of an increasingly experienced subject pool was explored by analyzing a series of experiments conducted over 2 years using the online labor market Amazon Mechanical Turk [Rand, D.G. *et al.* (2013) Intuitive cooperation and the social heuristics hypothesis: evidence from 15 time constraint studies, <http://ssrn.com/abstract=2222683>]. During that period, behavioral experiments became dramatically more common on Mechanical Turk, resulting in a subject pool that is highly experienced in study participation. As predicted by the social heuristics hypothesis, decisions made under time pressure became steadily less cooperative (as intuitions were eroded), whereas reflective responses remained constant. These findings suggest that intuitions are malleable rather than hard-coded. Thus, we find support for the social heuristics hypothesis, and for the importance of learning and culture in human cooperation.

Concluding remarks

Understanding the evolutionary dynamics of cooperation has important implications for our conceptualization of ourselves as human beings. Research in this field helps to explain the widespread cooperation that is a cornerstone of our existence as a supremely social species. It also

provides concrete guidance for individuals, organizations, and policy-makers seeking to promote cooperation in settings in which it is currently lacking.

In this review, we shed light on human cooperation by synthesizing theoretical research on evolutionary dynamics with experiments examining human behavior. We provide empirical evidence for five mechanisms for the evolution of human cooperation: direct reciprocity, indirect reciprocity, spatial selection, multilevel selection, and kin selection. We also highlight areas in which theory and experiments diverge, and for which more empirical and theoretical work is needed.

We then consider cooperation in one-shot anonymous settings in which no mechanisms are explicitly present. We provide evidence that cooperative strategies developed in the context of reciprocal interactions spill over into one-shot games. We show that this inclination towards intuitive reciprocation is malleable. Together, these results highlight the importance of reciprocity for human cooperation, as well as the powerful role played by learning and culture. The evidence we present does not rule out the possibility that (i) some level of one-shot cooperation was specifically favored by selection or that (ii) genetic evolution played an important role in the evolution of human cooperation. Conducting experiments to further distinguish between these hypotheses for the origins of human

Box 6. Questions for future research

- In the context of indirect reciprocity, what incentive is there to provide honest information about the actions of others? Why are people willing to risk their own reputation by defecting against those who are in bad standing?
- How is the role of population structure in human cooperation best assessed experimentally? There is overwhelming theoretical evidence that population structure affects evolutionary outcomes, and that some structures and update rules promote cooperation. However, several laboratory experiments on human subjects in fixed spatial arrangements do not find increased cooperation. Exploring which experimental settings might reveal strong effects of structure on cooperation is an important direction for future work.
- To what extent is cooperation in one-shot anonymous settings explained by the overgeneralization of cooperative strategies from reciprocal interactions? What experiments best discriminate between different evolutionary hypotheses for the origins of human cooperation?
- How does evidence of mechanisms for the evolution of cooperation in laboratory experiments generalize to real-world field settings? How can these mechanisms be most effectively harnessed to increase cooperation and promote social welfare in our societies?
- How will the addition of psychological complexity affect evolutionary game theory? The simplicity of the strategy space of most evolutionary models does not reflect the intricacies of human psychology and decision-making. Adding more psychological details to these models can generate great insights into human cooperation. For example, empirical evidence suggests that the brain is best represented not by a single agent but by multiple agents that are in conflict with each other. Such conflict between different motivations and cognitive processes should be incorporated into evolutionary models.
- Evolutionary models are typically focused on actions. Reciprocal strategies respond to the actions of others. However, a key element of human cooperation is assessing the intentions of others and modifying responses accordingly. Evolutionary models that explore the interplay between intentions and outcomes will provide powerful insights into human cooperation and morality.

cooperation in one-shot interactions is a fundamental challenge for the field.

Evolutionary game theoretic models can play a central role in this endeavor. Typically, these models treat agents as psychological black boxes: each agent has a particular hardcoded behavioral strategy, and selection operates on those strategies. Psychological complexity can, however, be added to evolutionary models (Box 6). Doing so will allow the generation of formal, testable predictions about the psychology produced under different evolutionary scenarios. Thus, evolutionary models with psychological complexity will link ultimate causation to the proximate psychology that can be assessed experimentally. Creating models in which different hypotheses about the evolution of human cooperation make divergent predictions about the resulting psychology will be invaluable for resolving this debate.

Critically, all of the perspectives on the evolution of human cooperation outlined in this review share a central message: selective forces from outside the laboratory influence play inside, effecting behavior in one-shot anonymous games. This key insight is often overlooked. Behavior in the laboratory cannot be explained without considering the environment in which that behavior evolved.

Acknowledgments

We thank Coren Apicella, Samuel Bowles, Nicholas Christakis, Anna Dreber, Jillian Jordan, Gordon Kraft-Todd and Peter Richerson for helpful feedback and comments on earlier drafts. Funding from the John Templeton Foundation is gratefully acknowledged.

References

- Nowak, M.A. (2006) Five rules for the evolution of cooperation. *Science* 314, 1560–1563
- Trivers, R. (1971) The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57
- Fudenberg, D. and Maskin, E. (1986) The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54, 533–554
- Binmore, K. and Samuelson, L. (1992) Evolutionary stability in repeated games played by finite automata. *J. Econ. Theory* 57, 278–305
- Sigmund, K. (2010) *The Calculus of Selfishness*. Princeton University Press
- Axelrod, R. (1984) *The Evolution of Cooperation*. Basic Books
- Nowak, M.A. and Sigmund, K. (1994) The alternating prisoner's dilemma. *J. Theor. Biol.* 168, 219–226
- Nowak, M. and Sigmund, K. (1993) A strategy of win–stay, lose–shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* 364, 56–58
- Nowak, M.A. and Sigmund, K. (1992) Tit for tat in heterogeneous populations. *Nature* 355, 250–253
- Nowak, M.A. and Sigmund, K. (1998) Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577
- Nowak, M.A. and Sigmund, K. (2005) Evolution of indirect reciprocity. *Nature* 437, 1291–1298
- Ohtsuki, H. and Iwasa, Y. (2006) The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* 239, 435–444
- Ohtsuki, H. et al. (2009) Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* 457, 79–82
- Brandt, H. and Sigmund, K. (2006) The good, the bad and the discriminator – errors in direct and indirect reciprocity. *J. Theor. Biol.* 239, 183–194
- Nowak, M.A. and May, R.M. (1992) Evolutionary games and spatial chaos. *Nature* 359, 826–829
- Nowak, M.A. et al. (2010) Evolutionary dynamics in structured populations. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 19–30
- Tarnita, C.E. et al. (2011) Multiple strategies in structured populations. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2334–2337
- Tarnita, C.E. et al. (2009) Strategy selection in structured populations. *J. Theor. Biol.* 259, 570
- Hauert, C. and Doebeli, M. (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428, 643–646
- Hauert, C. and Imhof, L.A. (2012) Evolutionary games in deme structured, finite populations. *J. Theor. Biol.* 299, 106–112
- Skyrms, B. and Pemantle, R. (2000) A dynamic model of social network formation. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9340
- Lieberman, E. et al. (2005) Evolutionary dynamics on graphs. *Nature* 433, 312–316
- Ohtsuki, H. et al. (2006) A simple rule for the evolution of cooperation on graphs and social networks. *Nature* 441, 502–505
- Szabo, G. and Fath, G. (2007) Evolutionary games on graphs. *Phys. Rep.* 446, 97–216
- Santos, F.C. et al. (2006) Cooperation prevails when individuals adjust their social ties. *PLoS Comput. Biol.* 2, e140
- Fu, F. et al. (2008) Reputation-based partner choice promotes cooperation in social networks. *Phys. Rev. E* 78, 026117
- Tarnita, C.E. et al. (2009) Evolutionary dynamics in set structured populations. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8601–8604
- Perc, M. and Szolnoki, A. (2010) Coevolutionary games – a mini review. *Biosystems* 99, 109–125
- Cavaliere, M. et al. (2012) Prosperity is associated with instability in dynamical networks. *J. Theor. Biol.* 299, 126–138
- Antal, T. et al. (2009) Evolution of cooperation by phenotypic similarity. *Proc. Natl. Acad. Sci. U.S.A.* 106, 8597–8600
- Riolo, R.L. et al. (2001) Evolution of cooperation without reciprocity. *Nature* 414, 441–443
- Traulsen, A. and Schuster, H.G. (2003) Minimal model for tag-based cooperation. *Phys. Rev. E* 68, 046129
- Wilson, D.S. (1975) A theory of group selection. *Proc. Natl. Acad. Sci. U.S.A.* 72, 143–146
- Boyd, R. and Richerson, P. (1990) Group selection among alternative evolutionarily stable strategies. *J. Theor. Biol.* 145, 331–342
- Traulsen, A. and Nowak, M.A. (2006) Evolution of cooperation by multilevel selection. *Proc. Natl. Acad. Sci. U.S.A.* 103, 10952–10955
- Sober, E. and Wilson, D.S. (1998) *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press
- Bowles, S. (2009) Did warfare among ancestral hunter-gatherers affect the evolution of human social behaviors? *Science* 324, 1293–1298
- Bowles, S. and Gintis, H. (2011) *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton University Press
- Boyd, R. et al. (2003) The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 3531–3535
- Darwin, C. (1871) *The Descent of Man and Selection in Relation to Sex*. Murray
- Nowak, M.A. and Highfield, R. (2011) *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Free Press
- Hamilton, W.D. (1964) The genetical evolution of social behaviour. I. *J. Theor. Biol.* 7, 1–16
- Nowak, M.A. et al. (2010) The evolution of eusociality. *Nature* 466, 1057–1062
- Abbot, P. et al. (2011) Inclusive fitness theory and eusociality. *Nature* 471, E1–E4
- Nowak, M.A. et al. (2011) Nowak et al. reply. *Nature* 471, E9–E10
- van Veelen, M. et al. (2012) Direct reciprocity in structured populations. *Proc. Natl. Acad. Sci. U.S.A.* 109, 9929–9934
- Rand, D.G. (2012) The promise of Mechanical Turk: how online labor markets can help theorists run behavioral experiments. *J. Theor. Biol.* 299, 172–179
- Rapoport, A. and Chammah, A.M. (1965) *Prisoner's Dilemma: A Study in Conflict and Cooperation*. University of Michigan Press
- Roth, A.E. and Murnighan, J.K. (1978) Equilibrium behavior and repeated play of the prisoner's dilemma. *J. Math. Psychol.* 17, 189–198
- Murnighan, J.K. and Roth, A.E. (1983) Expecting continued play in prisoner's dilemma games: a test of several models. *J. Conflict Resol.* 27, 279–300
- Duffy, J. and Ochs, J. (2009) Cooperative behavior and the frequency of social interaction. *Games Econ. Behav.* 66, 785–812

- 52 Dal Bó, P. and Fréchet, G.R. (2011) The evolution of cooperation in infinitely repeated games: experimental evidence. *Am. Econ. Rev.* 101, 411–429
- 53 Dal Bó, P. (2005) Cooperation under the shadow of the future: experimental evidence from infinitely repeated games. *Am. Econ. Rev.* 95, 1591–1604
- 54 Dreber, A. *et al.* (2008) Winners don't punish. *Nature* 452, 348–351
- 55 Fudenberg, D. *et al.* (2012) Slow to anger and fast to forgive: cooperation in an uncertain world. *Am. Econ. Rev.* 102, 720–749
- 56 Fudenberg, D. and Maskin, E.S. (1990) Evolution and cooperation in noisy repeated games. *Am. Econ. Rev.* 80, 274–279
- 57 Nowak, M.A. *et al.* (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646–650
- 58 Rand, D.G. *et al.* (2013) Evolution of fairness in the one-shot anonymous ultimatum game. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2581–2586
- 59 Rand, D.G. and Nowak, M.A. (2012) Evolutionary dynamics in finite populations can explain the full range of cooperative behaviors observed in the centipede game. *J. Theor. Biol.* 300, 212–221
- 60 Manapat, M.L. *et al.* (2012) Stochastic evolutionary dynamics resolve the traveler's dilemma. *J. Theor. Biol.* 303, 119–127
- 61 Rand, D.G. *et al.* (2009) Direct reciprocity with costly punishment: generous tit-for-tat prevails. *J. Theor. Biol.* 256, 45–57
- 62 Levin, S.A. (2000) *Fragile Dominion: Complexity and the Commons*. Basic Books
- 63 Hardin, G. (1968) The tragedy of the commons. *Science* 162, 1243–1248
- 64 Boyd, R. and Richerson, P.J. (1988) The evolution of reciprocity in sizable groups. *J. Theor. Biol.* 132, 337–356
- 65 Ostrom, E. *et al.* (1992) Covenants with and without a sword: self-governance is possible. *Am. Polit. Sci. Rev.* 86, 404–417
- 66 Fehr, E. and Gächter, S. (2000) Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* 90, 980–994
- 67 Rand, D.G. *et al.* (2009) Positive interactions promote public cooperation. *Science* 325, 1272–1275
- 68 Sefton, M. *et al.* (2007) The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* 45, 671–690
- 69 Sutter, M. *et al.* (2010) Choosing the stick or the carrot? Endogenous institutional choice in social dilemma situations. *Rev. Econ. Stud.* 77, 1540–1566
- 70 Choi, J.-K. and Ahn, T.K. (2013) Strategic reward and altruistic punishment support cooperation in a public goods game experiment. *J. Econ. Psychol.* 37, 17–30
- 71 Nikiforakis, N. (2008) Punishment and counter-punishment in public goods games: can we still govern ourselves? *J. Public Econ.* 92, 91–112
- 72 Gächter, S. and Herrmann, B. (2011) The limits of self-governance when cooperators get punished: experimental evidence from urban and rural Russia. *Eur. Econ. Rev.* 55, 193–210
- 73 Gächter, S. and Herrmann, B. (2009) Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philos. Trans. R. Soc. B: Biol. Sci.* 364, 791–806
- 74 Herrmann, B. *et al.* (2008) Antisocial punishment across societies. *Science* 319, 1362–1367
- 75 Rand, D.G. and Nowak, M.A. (2011) The evolution of antisocial punishment in optional public goods games. *Nat. Commun.* 2, 434
- 76 Rand, D.G. *et al.* (2010) Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* 265, 624–632
- 77 Powers, S.T. *et al.* (2012) Punishment can promote defection in group-structured populations. *J. Theor. Biol.* 311, 107–116
- 78 García, J. and Traulsen, A. (2012) Leaving the loners alone: evolution of cooperation in the presence of antisocial punishment. *J. Theor. Biol.* 307, 168–173
- 79 Panchanathan, K. and Boyd, R. (2004) Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* 432, 499–502
- 80 Wedekind, C. and Milinski, M. (2000) Cooperation through image scoring in humans. *Science* 288, 850–852
- 81 Milinski, M. *et al.* (2002) Reputation helps solve the 'tragedy of the commons'. *Nature* 415, 424–426
- 82 Pfeiffer, T. *et al.* (2012) The value of reputation. *J. R. Soc. Interface* 9, 2791–2797
- 83 Semmann, D. *et al.* (2005) Reputation is valuable within and outside one's own social group. *Behav. Ecol. Sociobiol.* 57, 611–616
- 84 Rockenbach, B. and Milinski, M. (2006) The efficient interaction of indirect reciprocity and costly punishment. *Nature* 444, 718–723
- 85 Sommerfeld, R.D. *et al.* (2007) Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17435–17440
- 86 Jacquet, J. *et al.* (2011) Shame and honour drive cooperation. *Biol. Lett.* 7, 899–901
- 87 Ule, A. *et al.* (2009) Indirect punishment and generosity toward strangers. *Science* 326, 1701–1704
- 88 Seinen, I. and Schram, A. (2006) Social status and group norms: indirect reciprocity in a repeated helping experiment. *Eur. Econ. Rev.* 50, 581–602
- 89 Bolton, G.E. *et al.* (2005) Cooperation among strangers with limited information about reputation. *J. Public Econ.* 89, 1457–1468
- 90 Lacetera, N. and Macis, M. (2010) Social image concerns and prosocial behavior: field evidence from a nonlinear incentive scheme. *J. Econ. Behav. Organ.* 76, 225–237
- 91 Karlan, D. and McConnell, M.A. (2012) *Hey Look at Me: The Effect of Giving Circles on Giving*. National Bureau of Economic Research
- 92 Yoeli, E. *et al.* (2013) Powering up with indirect reciprocity in a large-scale field experiment. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10424–10429
- 93 Dunbar, R.I.M. *et al.* (1997) Human conversational behavior. *Hum. Nat.* 8, 231–246
- 94 Resnick, P. *et al.* (2006) The value of reputation on eBay: a controlled experiment. *Exp. Econ.* 9, 79–101
- 95 Suzuki, S. and Kimura, H. (2013) Indirect reciprocity is sensitive to costs of information transfer. *Sci. Rep.* 3, 1435
- 96 Pahlia, N. *et al.* (2009) Dirty work, clean hands: the moral psychology of indirect agency. *Organ. Behav. Hum. Decis. Process.* 109, 134–141
- 97 Coffman, L.C. (2011) Intermediation reduces punishment (and reward). *Am. Econ. J. Microecon.* 3, 77–106
- 98 Hamlin, J.K. *et al.* (2007) Social evaluation by preverbal infants. *Nature* 450, 557–559
- 99 Hamlin, J.K. *et al.* (2011) How infants and toddlers react to antisocial others. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19931–19936
- 100 Anderson, J.R. *et al.* (2013) Third-party social evaluation of humans by monkeys. *Nat. Commun.* 4, 1561
- 101 Milinski, M. and Rockenbach, B. (2007) Spying on others evolves. *Science* 317, 464–465
- 102 Burnham, T. and Hare, B. (2007) Engineering human cooperation. *Hum. Nat.* 18, 88–108
- 103 Haley, K.J. and Fessler, D.M.T. (2005) Nobody's watching? Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* 26, 245–256
- 104 Hoffman, E. *et al.* (1996) Social distance and other-regarding behavior in dictator games. *Am. Econ. Rev.* 86, 653–660
- 105 Gore, J. *et al.* (2009) Snowdrift game dynamics and facultative cheating in yeast. *Nature* 459, 253–256
- 106 Grujić, J. *et al.* (2012) Consistent strategy updating in spatial and non-spatial behavioral experiments does not promote cooperation in social networks. *PLoS ONE* 7, e47718
- 107 Grujić, J. *et al.* (2010) Social experiments in the mesoscale: humans playing a spatial prisoner's dilemma. *PLoS ONE* 5, e13749
- 108 Traulsen, A. *et al.* (2010) Human strategy updating in evolutionary games. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2962–2966
- 109 Suri, S. and Watts, D.J. (2011) Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE* 6, e16836
- 110 Gracia-Lázaro, C. *et al.* (2012) Heterogeneous networks do not promote cooperation when humans play a prisoner's dilemma. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12922–12926
- 111 Allen, B. *et al.* (2011) How mutation affects evolutionary games on graphs. *J. Theor. Biol.* 299, 97–105
- 112 Semmann, D. (2012) Conditional cooperation can hinder network reciprocity. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12846–12847
- 113 Rand, D.G. *et al.* (2011) Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19193–19198
- 114 Wang, J. *et al.* (2012) Cooperation and assortativity with dynamic partner updating. *Proc. Natl. Acad. Sci. U.S.A.* 109, 14363–14368
- 115 Fehl, K. *et al.* (2011) Co-evolution of behaviour and social network structure promotes human cooperation. *Ecol. Lett.* 14, 546–551

- 116 Jordan, J.J. *et al.* (2013) Contagion of cooperation in static and fluid social networks. *PLoS ONE* 8, e66199
- 117 Efferson, C. *et al.* (2008) The coevolution of cultural groups and ingroup favoritism. *Science* 321, 1844–1849
- 118 Tajfel, H. *et al.* (1971) Social categorization and intergroup behavior. *Eur. J. Soc. Psychol.* 1, 149–178
- 119 Rand, D.G. *et al.* (2009) Dynamic remodeling of in-group bias during the 2008 presidential election. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6187–6191
- 120 Sherif, M. *et al.* (1961) *Intergroup Conflict and Cooperation: The Robbers' Cave Experiment*. Institute of Group Relations, University of Oklahoma
- 121 Marlowe, F.W. (2005) Hunter-gatherers and human evolution. *Evol. Anthropol.* 14, 54–67
- 122 Fu, F. *et al.* (2012) Evolution of in-group favoritism. *Sci. Rep.* 2, 460
- 123 Apicella, C.L. *et al.* (2012) Social networks and cooperation in hunter-gatherers. *Nature* 481, 497–501
- 124 O'Malley, A.J. *et al.* (2012) Egocentric social network structure, health, and pro-social behaviors in a national panel study of Americans. *PLoS ONE* 7, e36250
- 125 Fowler, J.H. *et al.* (2011) Correlated genotypes in friendship networks. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1993–1997
- 126 Puurtinen, M. and Mappes, T. (2009) Between-group competition and human cooperation. *Proc. R. Soc. B: Biol. Sci.* 276, 355–360
- 127 Erev, I. *et al.* (1993) Constructive intergroup competition as a solution to the free rider problem: a field experiment. *J. Exp. Soc. Psychol.* 29, 463–478
- 128 Gunnthorsdottir, A. and Rapoport, A. (2006) Embedding social dilemmas in intergroup competition reduces free-riding. *Organ. Behav. Hum. Decis. Process.* 101, 184–199
- 129 Bornstein, G. *et al.* (1990) Intergroup competition as a structural solution to social dilemmas. *Soc. Behav.* 5, 247–260
- 130 Tan, J.H.W. and Bolle, F. (2007) Team competition and the public goods game. *Econ. Lett.* 96, 133–139
- 131 Sääksvuori, L. *et al.* (2011) Costly punishment prevails in intergroup conflict. *Proc. R. Soc. B: Biol. Sci.* 278, 3428–3436
- 132 Böhm, R. and Rockenbach, B. (2013) The inter-group comparison – intra-group cooperation hypothesis: comparisons between groups increase efficiency in public goods provision. *PLoS ONE* 8, e56152
- 133 Voors, M.J. *et al.* (2012) Violent conflict and behavior: a field experiment in Burundi. *Am. Econ. Rev.* 102, 941–964
- 134 Gneezy, A. and Fessler, D.M.T. (2011) Conflict, sticks and carrots: war increases prosocial punishments and rewards. *Proc. R. Soc. B: Biol. Sci.* 279, 219–223
- 135 Williams, G.C. (1966) *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press
- 136 Bell, A.V. *et al.* (2009) Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17671–17674
- 137 Lieberman, D. *et al.* (2007) The architecture of human kin detection. *Nature* 445, 727–731
- 138 Camerer, C.F. (2003) *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press
- 139 Choi, J.K. and Bowles, S. (2007) The coevolution of parochial altruism and war. *Science* 318, 636–640
- 140 Delton, A.W. *et al.* (2011) Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl. Acad. Sci. U.S.A.* 108, 13335–13340
- 141 Richerson, P.J. and Boyd, R. (2005) *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press
- 142 Boyd, R. and Richerson, P.J. (1982) Cultural transmission and the evolution of cooperative behavior. *Hum. Ecol.* 10, 325–351
- 143 Chudek, M. and Henrich, J. (2011) Culture gene coevolution, norm-psychology and the emergence of human prosociality. *Trends Cogn. Sci.* 15, 218–226
- 144 Bowles, S. *et al.* (2003) The co-evolution of individual behaviors and social institutions. *J. Theor. Biol.* 223, 135–147
- 145 Cosmides, L. and Tooby, J. (2005) Neurocognitive adaptations designed for social exchange. In *The Handbook of Evolutionary Psychology* (Buss, D.M., ed.), pp. 584–627, John Wiley & Sons
- 146 Rand, D.G. *et al.* (2012) Spontaneous giving and calculated greed. *Nature* 489, 427–430
- 147 Rand, D.G. *et al.* (2013) Rand *et al.* reply. *Nature* 497, E1–E2
- 148 Roch, S.G. *et al.* (2000) Cognitive load and the equality heuristic: a two-stage model of resource overconsumption in small groups. *Organ. Behav. Hum. Decis. Process.* 83, 185–212
- 149 Schulz, J.F. *et al.* (2012) Affect and fairness: dictator games under cognitive load. *J. Econ. Psychol.* <http://dx.doi.org/10.1016/j.joep.2012.08.007>
- 150 Cornelissen, G. *et al.* (2011) Are social value orientations expressed automatically? Decision making in the dictator game. *Pers. Soc. Psychol. Bull.* 37, 1080–1090
- 151 Bartlett, M.Y. and DeSteno, D. (2006) Gratitude and prosocial behavior: helping when it costs you. *Psychol. Sci.* 17, 319–325
- 152 DeSteno, D. *et al.* (2010) Gratitude as moral sentiment: emotion-guided cooperation in economic exchange. *Emotion* 10, 289
- 153 DeSteno, D. (2009) Social emotions and intertemporal choice 'hot' mechanisms for building social and economic capital. *Curr. Dir. Psychol. Sci.* 18, 280–284
- 154 Sanfey, A.G. *et al.* (2003) The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755–1758
- 155 Sutter, M. *et al.* (2003) Bargaining under time pressure in an experimental ultimatum game. *Econ. Lett.* 81, 341–347
- 156 Gospic, K. *et al.* (2011) Limbic justice – amygdala involvement in immediate rejection in the ultimatum game. *PLoS Biol.* 9, e1001054
- 157 Grimm, V. and Mengel, F. (2011) Let me sleep on it: delay reduces rejection rates in ultimatum games. *Econ. Lett.* 111, 113–115
- 158 Knoch, D. *et al.* (2006) Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science* 314, 829–832
- 159 van't Wout, M. *et al.* (2005) Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport* 16, 1849–1852
- 160 Nowak, M.A. *et al.* (2000) Fairness versus reason in the ultimatum game. *Science* 289, 1773–1775
- 161 Hauge, K.E. *et al.* (2009) Are social preferences skin deep? Dictators under cognitive load. In *Working Paper in Economics*, University of Gothenburg
- 162 Tinghög, G. *et al.* (2013) Intuition and cooperation reconsidered. *Nature* 497, E1–E2
- 163 Hauert, C. *et al.* (2006) Synergy and discounting of cooperation in social dilemmas. *J. Theor. Biol.* 239, 195
- 164 Nowak, M.A. (2012) Evolving cooperation. *J. Theor. Biol.* 299, 1–8
- 165 Archetti, M. (2009) The volunteer's dilemma and the optimal size of a social group. *J. Theor. Biol.* 261, 475–480
- 166 Nowak, M.A. and Roch, S. (2007) Upstream reciprocity and the evolution of gratitude. *Proc. R. Soc. B: Biol. Sci.* 274, 605–610
- 167 Rankin, D.J. and Taborsky, M. (2009) Assortment and the evolution of generalized reciprocity. *Evolution* 63, 1913–1922
- 168 Nakamaru, M. and Iwasa, Y. (2005) The evolution of altruism by costly punishment in lattice-structured populations: score-dependent viability versus score-dependent fertility. *Evol. Ecol. Res.* 7, 853–870
- 169 García, J. and van den Bergh, J.C.J.M. (2011) Evolution of parochial altruism by multilevel selection. *Evol. Hum. Behav.* 32, 277–287
- 170 Masuda, N. (2012) Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. *J. Theor. Biol.* 311, 8–18
- 171 Hammond, R.A. and Axelrod, R. (2006) The evolution of ethnocentrism. *J. Conflict Resol.* 50, 926–936
- 172 Fehr, E. and Fischbacher, U. (2004) Third-party punishment and social norms. *Evol. Hum. Behav.* 25, 63–87
- 173 Fehr, E. and Gächter, S. (2002) Altruistic punishment in humans. *Nature* 415, 137–140
- 174 Yamagishi, T. (1986) The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* 51, 110–116
- 175 Almenberg, J. *et al.* (2011) Third party reward and punishment: group size, efficiency and public goods. In *Psychology and Punishment* (Palmetti, N.M. and Russo, J.P., eds), pp. 73–92, Nova Science Publishers
- 176 Janssen, M.A. and Bushman, C. (2008) Evolution of cooperation and altruistic punishment when retaliation is possible. *J. Theor. Biol.* 254, 541–545
- 177 Brewer, M.B. (1979) In-group bias in the minimal intergroup situation: a cognitive-motivational analysis. *Psychol. Bull.* 86, 307
- 178 Fowler, J.H. and Kam, C.D. (2007) Beyond the self: social identity, altruism, and political participation. *J. Polit.* 69, 813–827