Cooperation is the third fundamental principle of evolution, beside mutation and selection. Mutation generates variation upon which selection acts. Cooperation leads to the integration of competing units and allows the formation of higher levels of organization. The emergence of the first cells, of eucarya, of multi-cellular organisms, of animal societies, and of humankind are the consequences of cooperation. Cooperation is the creative force of evolution, which allows the emergence of complexity and structure and ultimately human life. Cooperation is never fully stable, but waxes and wanes. Natural selection tends to oppose cooperation unless specific mechanisms for evolution of cooperation are operative in a population (Nowak 2006). Presently five such mechanisms are known: direct reciprocity, indirect reciprocity, spatial selection, group selection, and kin selection. This chapter will discuss theoretical and empirical evidence for each of those mechanisms (Rand and Nowak 2013).

The importance and fragility of cooperation is evident in human interactions. We attempt to cooperate and to realize the potential benefits of cooperation, but the temptation to defect is always present. Humans are ‘SuperCooperators’ (Nowak and Highfield 2011) in the sense that they can utilize and know all five mechanisms for evolution of cooperation. Knowledge of this research endeavor, which exists at the interface of mathematics, evolutionary biology, psychology, and economics, is especially relevant in situations that see a breakdown of human cooperation at the global level and at a time when cooperation would be most needed to solve the world’s pressing problems (Messner and Weinlich in this volume).

In a cooperative (or social) dilemma, there is a 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. A more detailed definition can be provided by game theory. Consider a game
between two strategies, C and D, and the payoff matrix (indicating the row player’s payoff):

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>R</td>
<td>S</td>
</tr>
<tr>
<td>D</td>
<td>T</td>
<td>P</td>
</tr>
</tbody>
</table>

When does it make sense to call strategy C ‘cooperation’ and strategy D ‘defection’? The following definition (Hauert et al. 2006: 195; Nowak 2012) is useful: the game is a cooperative dilemma if (1) two cooperators get a higher payoff than two defectors, R > P, yet (2) there is an incentive to defect. This incentive can arise in three different ways: (2a) if T > R, then it is better to defect when playing against a cooperator; (2b) if P > S, then it is better to defect when playing against a defector; and (2c) 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 those three conditions holds, then we have a cooperative dilemma. If none hold, 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 (Axelrod 1984; Rapoport and Chammah 1965).

The Prisoner’s Dilemma (PD) is the most stringent cooperative dilemma. Here defectors dominate cooperators. In a well-mixed population natural selection always favors defectors over cooperators. For cooperation to arise in the PD we need a mechanism for the evolution of cooperation. Cooperative dilemmas which are not the PD 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.

There are a few other conditions that one could consider. For example, it might be natural to demand that it is always better to meet a cooperator than a defector (regardless of your own strategy). This implies R > S and T > P. Moreover, if 2R > T + S, then the total payoff of the population is maximized if everyone cooperates; otherwise a mixed population achieves the highest total payoff.

The definition of a cooperative dilemma can be generalized to more than two people, which brings us to the theory of n person games. Denote the payoffs for cooperators and defectors by $P_i$ and $Q_{ip}$, respectively, in groups that contain $i$ cooperators and $n - i$ defectors. For the game to be a cooperative dilemma we require that (1) an all-cooperator group gets a higher payoff than an all-defector group, $P_n > Q_0$, yet (2) there is some incentive to defect. The incentives to defect can take the following form: (2a) $P_i < Q_{i+1}$ for $i = 1 \ldots n$ and (2b) $P_i < Q_i$ for $i = 1 \ldots n - 1$. The conditions (2a) mean that an individual can increase its payoff by switching from cooperation to defection. The conditions (2b) mean that in any mixed group, defectors have a higher payoff than cooperators. If only some of these incentives (2) hold, then we have a relaxed cooperative dilemma. In this case some evolution of cooperation is possible even without a specific mechanism. But 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 (Archetti 2009). If all incentives hold, we have the n person equivalent of a Prisoner’s Dilemma, called the ‘Public Goods Game’ (PGG) (Hardin 1968), and a mechanism for evolution of cooperation is needed.

In a well-mixed population, where each individual is equally likely to interact and compete with each other, individual, natural selection favors defection in the PD. 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 (Nowak 2006). 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: direct reciprocity, indirect reciprocity, spatial selection, multi-level selection and kin selection.

It is important to distinguish between interaction patterns that are mechanisms for the evolution of cooperation and behaviors that are not mechanisms but instead require an evolutionary explanation themselves. 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 rewarding cooperation and punishing 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 explain the evolution of cooperation: adding one or more of these elements to a PD will not cause selection to favor cooperation. Instead, these concepts are descriptions of behavior which require an evolutionary explanation. Group selection, spatial structure, or some chance of direct or indirect reciprocity can lead to the evolution of upstream reciprocity (Nowak and Roch 2007; Rankin and Taborsky 2009), strong reciprocity (Boyd et al. 2003; Nakamaru and Iwasa 2005; Ohtsuki et al. 2009), and parochial altruism (Choi and Bowles 2007; Fu et al. 2012; García and van den Bergh 2011; Hammond and Axelrod 2006; Masuda 2012).

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

Theoretical basis of the five mechanisms

(1) Direct reciprocity arises if there are repeated encounters between the same two individuals (Binmore and Samuelson 1992; Fudenberg and Maskin 1986; Sigmund 2010; Trivers 1971). Because they interact repeatedly, individuals can use ‘conditional strategies’, where behavior depends on previous outcomes. Direct reciprocity allows the evolution of cooperation if the probability of another interaction is sufficiently high (Axelrod 1984). Under this ‘shadow of the future’ I may pay the cost of cooperation today in order to earn your reciprocal cooperation tomorrow. The repeated game can occur with players making simultaneous decisions in each round or taking turns (Nowak and Sigmund 1994). Successful strategies for the simultaneous repeated PD include tit for tat (TFT), generous tit for tat (Nowak and Sigmund 1992), and win-stay, lose-shift (Nowak and Sigmund 1993). 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 (Generous TFT) (Nowak and Sigmund 1992).

(2) 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 these encounters can spread through communication, affecting the reputations of the participants. Individuals can thus adopt conditional strategies that base their decisions on the reputation of the recipient (Nowak and Sigmund 1998, 2005). 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 (Brandt and Sigmund 2006; Ohtsuki and Iwasa 2006; Ohtsuki et al. 2009). The social norm specifies how reputations are updated based on 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. In the context of indirect reciprocity, increasing scale and complexity of interactions is beneficial for cooperation, as more interactions allow greater flow of reputational information.

(3) Spatial selection can favor cooperation without the need for strategic complexity (Nowak and May 1992; Nowak et al. 2010a). 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 which prevail even if surrounded by defectors. The fundamental idea is that clustering creates assortment, where 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 (Tarnita et al. 2009b; Tarnita et al. 2011). 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 (Hauert and Doebeli 2004; Hauert and Imhof 2012) or social networks (Skyrms and Pemantle 2000), and can be static (Lieberman et al. 2005; Ohtsuki et al. 2006; Szabo and Fath 2007) or dynamic (Cavaliere et al. 2012; Fu et al. 2008; Perc and Szolnoki 2010; Santos et al. 2006; Skyrms and Pemantle 2000; Tarnita et al. 2009a). Population structure can also be implemented through tag-based cooperation, where interaction and cooperation are determined by arbitrary tags or markers (Antal et al. 2009; Riolo et al. 2001; Traulsen and Schuster 2003). In this case, clustering is not literally spatial but instead occurs in the space of phenotypes (Antal et al. 2009). Models of spatial selection generally suggest that as the scale interaction and degree of interconnectedness increase, it becomes more difficult to maintain cooperation: interconnection undermines clustering.

(4) Multi-level selection operates if, in addition to competition between individuals in a group, there is also competition between groups (Bowles 2009; Bowles and Gintis 2011; Boyd et al. 2003; Boyd and Richerson 1990; Sober and Wilson 1998; Traulsen and Nowak 2006; Wilson 1975). 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.

(Darwin 1871: 159)

Interestingly, multi-level selection relies on competition between distinct groups, and thus may be challenged in the face of an increasingly interconnected and global world.

(5) 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. Much of the current kin selection literature, however, does not adhere to this simple definition based on kin recognition. Instead kin selection is linked to the concept of inclusive fitness (Hamilton 1964). 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 it makes strong assumptions that prevent it from being a general concept (Nowak et al. 2010b). 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, which is critical of inclusive
excellent job of organizing the experimental data across a number of repeated game experiments (Rand and Nowak 2013). This is one of numerous situations in which stochastic evolutionary game theory (Nowak et al. 2004) successfully describes observed human behavior (Manapat et al. 2012b; Rand and Nowak 2012; Rand et al. 2009a; Rand et al. 2013a).

Repetition promotes cooperation in dyadic interactions. The situation is more complicated if groups of players interact repeatedly (Levin 2000). Such group cooperation is studied in the context of the Public Goods Game (PGG) (Hardin 1968), 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 versus how much to contribute to the group. All contributions are multiplied by some constant $r$, with $1 < r < n$, 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 multi-player games, this stability is fragile and can be undermined by errors or a small fraction of defectors (Boyd and Richardson 1988): for example, if I know that others will only contribute next period so long as all players contribute this period, then I maximize my payoff by contributing; but as soon as one person accidentally defects (or decides to experiment with defection), all players switch to defection. As a result, cooperation almost always fails in repeated PGGs in the laboratory (Fehr and Gächter 2000; Ostrom et al. 1992; Rand et al. 2009c).

Does this mean that mechanisms other than direct reciprocity are needed to explain cooperation among groups? We argue that the answer is 'no'. One must only realize that group interactions do not occur in a vacuum, but rather are superimposed on a network of dyadic personal relationships. In these personal pairwise relationships, people can condition their behavior on the other's previous conduct in the group. This allows for the targeted reciprocity that is missing in the PGG, giving us the power to enforce group-level cooperation. Dyadic relationships can be represented by adding pairwise rewards or punishment opportunities to the PGG. After each PGG round, subjects can pay to increase or decrease the payoff of other group members based on their contributions: high contributors are typically rewarded, and low contributors punished (Ellingsen et al. 2012; Fehr and Gächter 2000, 2002; Rand et al. 2009c). 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 (Choi and Ahn 2013; Fehr and Gächter 2000; Ostrom et al. 1992; Rand et al. 2009c; Sefton et al. 2007; Sutter et al. 2010). Naturally, given that both implementations of direct reciprocity promote cooperation to an equal extent, higher payoffs are achieved when using reward (which creates benefit) than punishment
(which destroys it). Rewarding also avoids vendettas (Dreber et al. 2008; Niki- forakis 2008) and the possibility of ‘antisocial punishment’, where low contributors pay to punish high contributors. Antisocial punishment has been demonstrated to occur in cross-cultural laboratory experiments (Gächter and Herrmann 2009, 2011; Herrmann et al. 2008), and can prevent the evolution of cooperation in theoretical models (García and Traulsen 2012; Powers et al. 2012; Rand and Nowak 2011; Rand et al. 2010). These cross-cultural studies 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 et al. 2012). With repetition and the addition of pairwise interactions, cooperation can be a robust equilibrium in the PGG, but populations can nonetheless get stuck in other less efficient equilibria or fail to equilibrate at all (Ellingsen et al. 2012).

Taken together, the many experiments exploring the linking of dyadic and multi-player 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, Panchanathan and Boyd 2004). 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 lab, subjects typically play with randomly matched partners and are informed about these partners’ choices in previous interactions with others (Milinski et al. 2002; Wedekind and Milinski 2000). 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 (Bolton et al. 2005; Jacquet et al. 2011; Milinski et al. 2002; Pfeiffer et al. 2012; Rockenbach and Milinski 2006; Seinen and Schram 2006; Semmann et al. 2005; Sommerfeld et al. 2007; Ule et al. 2009; Wedekind and Milinski 2000). Thus having the reputation for being a cooperator is valuable, and cooperation is maintained: it is worth paying the cost of cooperation today in order to earn the benefits of a good reputation tomorrow.

Reputation effects have also been shown to promote pro-social behavior outside of the laboratory. Field experiments, conducted in real-world settings with participants that do not know they are part of an experiment, find that publicizing the names of donors increases the level of blood donation (Lacetera and Macis 2010) and giving to charity (Karlan and McConnell 2012). Non-financial incentives involving reputation have also been shown to outperform monetary incentives in motivating participation in an energy blackout prevention program in
California (Yoeli et al. 2013) and the sale of condoms on behalf of a health organization in Namibia (Ashraf et al. 2012).

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 (Dunbar et al. 1997; Sommerfeld et al. 2007). 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 (Resnick et al. 2006). 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 2011).

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 (Suzuki and Kimura 2013). 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. Yet doing so can potentially be damaging for your own reputation. This is particularly true when using simple reputation systems such as image scoring (Nowak and Sigmund 1998), which is a first order assessment rule evaluating actions only (cooperation is good, defection is bad). But it can apply even when using more complex reputation rules, where 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 with bad reputations. Thus they avoid getting exploited, while also not having to defect themselves. Such a system should lead to stable cooperation using even the simplest of reputation systems. This solution brings indirect reciprocity together with theories of partner choice (Fu et al. 2008; Manapat et al. 2012a; Skyrms and Pemantle 2000). Another interesting possibility involves intermediation: if you employ an intermediary to defect against bad players on your behalf, this may help to avoid sullying your reputation. Consistent with this possibility, experimental evidence suggests that the use of intermediaries reduces blame for selfish actions (Coffmann 2011; Paharia et al. 2009). 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 for the central role of reputational concerns in human evolution. Infants as young as six months of age take into account others' actions toward third parties when making social evaluations (Hamlin et al. 2007; Hamlin et al. 2011). This tendency even occurs between species: capuchin monkeys are less likely to accept food from humans who were unhelpful to third parties (Anderson et al. 2013). Humans are also exquisitely sensitive to the possibility of being observed by third parties (Milinski and Rockenbach 2007). For example, people are even more pro-social when being watched by a robot with large fake eyes (Burnham and Hare 2007), or when a pair of stylized eye-spots are added to the computer's desktop background (Haley and Fessler 2005). In the opposite direction, making studies double-blind such that the experimenters cannot associate subjects with their actions increases selfishness (Hoffman et al. 1996).

**Spatial selection**

Unlike direct and indirect reciprocity, experimental evidence of spatial selection among humans is less clear. (There is good evidence for spatial selection in unicellular organisms; Gore et al. 2009.) Experiments that investigate fixed spatial structure typically find no increase in cooperation. These experiments assign subjects to locations in a network and have them play repeatedly with their neighbors. Cooperation rates are then compared to a control where subjects' positions in the network are randomly reshuffled each round, creating a well-mixed population. As in the 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 each round with the payoff of each neighbor, as well as that neighbor's choice. Yet unlike the models, cooperation rates in these experiments are no higher in structured populations than in randomly shuffled populations (Gracia-Lázaro et al. 2012; Grujić et al. 2012; Grujić et al. 2010; Suri and Watts 2011; Traulsen et al. 2010).

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 (Traulsen et al. 2010). 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 (Allen et al. 2011). Without sufficient clustering, cooperation is no longer advantageous.

Another explanation involves the way subjects choose which strategy to adopt. Theoretical models make detailed assumptions about how individuals update their strategies, and whether network structure can promote cooperation in these models depends critically on these details (Tarnita et al. 2009b). It is possible that human subjects in the experimental situations examined thus far tend to update rules that cancel the effect of spatial structure (Traulsen et al. 2010). A related argument involves the confounding of spatial structure and direct reciprocity that occurs in these experiments (Semmann 2012). 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 the tendency to reciprocate may lead to the demise of cooperation.

Here we offer a possible alternative explanation. Theoretical work has shown that cooperation is not always expected to succeed in structure populations. Instead, particular conditions are required. For example, under a particular set of assumptions about strategy updating, cooperation is only predicted to be favored when the PD’s benefit-to-cost ratio exceeds the average number of neighbors in the network (Ohtsuki et al. 2006). Thus it may be that previous experiments found no effects of networked interaction because they did not explore the right combinations of payoffs and network structures. Exploring this possibility is an important direction for future study.

In contrast to these negative results using static networks, dynamic networks robustly promote cooperation in the laboratory (Fehl et al. 2011; Jordan et al. 2013; Rand et al. 2011; Wang et al. 2012). In these experiments subjects can make or break connections with others, and the network evolves over time. This dynamic character allows subjects to engage in targeted action via ‘link reciprocity’: players can choose to offer links to defectors or to make links with cooperators. The importance of dynamic assortment based on arbitrary tags has also been demonstrated in lab experiments using coordination games: associations between tags and actions emerge spontaneously, as does preferential interaction between players sharing the same tag (Efferson et al. 2008).

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 (Tajfel et al. 1971). Alternatively, the introduction of a higher-level threat can realign coalitions, making yesterday’s enemies into today’s allies (Rand et al. 2009b; Sherif et al. 1961). Such plasticity is not limited to modern humans: many early human societies were characterized by fission-fusion dynamics, where group membership changed regularly (Marlowe 2005). Developing 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 (Fu et al. 2012; Tarnita et al. 2009a) and tag-based cooperation (Antal et al. 2009; Riolo et al. 2001; Traulsen and Schuster 2003) represent steps in this direction.

Finally, studies examining behavior in real-world networks also provide evidence for the importance of population structure in cooperation. For example, experiments with hunter-gatherers show that social ties predict similarity in cooperative behavior (Apicella et al. 2012). A nationally representative survey of American adults found that people who engage in more pro-social behavior have more social contacts, as predicted by dynamic network models (O’Malley et al. 2012). There is also evidence that social structure is heritable (Fowler et al. 2011), 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.

**Multi-level selection**

In the laboratory, multi-level selection is typically implemented using interaction structures where groups compete over resources. 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 defection is typically still the payoff-maximizing choice. Numerous such experiments have shown that competition between groups increases cooperation substantially (Bornstein et al. 1990; Erev et al. 1993; Gunnthorsdottir and Rapoport 2006; Puurtinen and Mappes 2009; Sääksvuori et al. 2011; Tan and Bolle 2007). Furthermore, just phrasing the interaction as a competition between groups, without any monetary prize for winning, also increases cooperation (Böhm and Rockenbach 2013; Tan and Bolle 2007). Experience with real-world inter-group conflict also increases cooperation (Gneezy and Fessler 2012; Voors et al. 2012). In sum, there is ample evidence that inter-group competition can be a powerful force for promoting within-group cooperation.

Critics of multi-level selection argue that, empirically, the conditions necessary for substantial selection pressure at the level of group were not met over the course of human history (Williams 1966): concerns include low ratios of between-group versus within-group variation due to factors such as migration, mutation/experimentation, and 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 in human history the conditions necessary for multi-level selection have been satisfied, either at the genetic (Bowles 2009; Bowles and Gintis 2011) or cultural level (Bell et al. 2009).

**Kin selection**

Kin selection is the least studied mechanism when it comes to 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 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, as 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 and kin selection in the context of human cooperation is important. It is essential to remember that 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) (Lieberman et al. 2007): 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 (for example, if you are a younger sibling that 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 as an indication of the probability of future interaction.

More studies on this topic are needed, in particular developing 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, thereby overcoming the limitations of inclusive fitness (Nowak et al. 2010b).

Cooperation in the absence of any mechanisms

How can we explain cooperation in one-shot anonymous interactions between strangers? Such cooperation is common (Bowles and Gintis 2011; Camerer 2003; Nowak and Highfield 2011; Rand and Nowak 2013; Tomasello 2009), yet it seems to contradict theoretical predictions because none of the five mechanisms appear 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 main dimensions along which scholars disagree: (1) whether cooperation in one-shot interactions was explicitly favored by evolution (through spatial or multi-level selection), or whether such altruistic cooperation is the result of overgeneralizing strategies from settings where cooperation is long-run self-interested (e.g. due to direct and indirect reciprocity); and (2) the relative importance of genetic evolution versus cultural evolution in shaping human cooperation.

On the first dimension, one perspective argues that multi-level selection and spatial structure specifically favored cooperation in one-shot anonymous settings (Bowles and Gintis 2011; Boyd et al. 2003; Choi and Bowles 2007). 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 altruistic preferences. 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. As cooperating is typically advantageous, we internalize cooperation as our default behavior. This cooperative predisposition is then sometimes overgeneralized, spilling over into the unusual situations where others are not watching (Delton et al. 2011; Haley and Fessler 2005). 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 who cooperate even when there are no future benefits from doing so: the disagreement is over whether or not one-shot cooperation was directly favored by selection or whether it is a byproduct of selection in non-anonymous interactions.

Turning to the second dimension, all of the mechanisms can function via either genetic evolution or cultural evolution. In the context of cultural evolution, traits spread through learning, often modeled as imitation of strategies which earn higher payoffs or are more common (Richerson and Boyd 2005). Multi-level selection has been argued by some to promote cooperation through genetic evolution (Sober and Wilson 1998), while others posit an important role of culture (Bowles and Gintis 2011; Bowles et al. 2003; Boyd and Richerson 1982; Chudek and Henrich 2011). The same is true of reciprocity: we might have genetic predispositions to cooperate because our ancestors lived in small groups with largely repeated interactions (Cosmides and Tooby 2005; Delton et al. 2011). 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 – the ‘Social Heuristics Hypothesis’ (Peyiekovich and Rand in press; Rand et al. 2012; Rand et al. 2013a). 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, while reflection and deliberation lead to selfishness. Inducing an intuitive mindset through priming or time pressure can increase cooperation relative to a more reflective mindset (Rand and Kraft-Todd 2014; Rand et al. 2012; Rand et al. 2013b; Rand et al. 2014). Increasing the role of intuition through cognitive load augments generosity in a resource allocation game (Roch et al. 2000) and in a unilateral money division task (i.e. dictator game; Cornelissen et al. 2011; Schulz et al. 2014). Affective, emotional responses play an important role in pro-social decision making (Bartlett and DeSteno 2006; DeSteno 2009; DeSteno et al. 2010). These findings suggest that cooperation in one-shot anonymous interactions involves some overgeneralization: intuitive, emotional processes favor the
typically advantageous behavior of cooperation, while reflection and reasoning adjust toward the behavior that is payoff maximizing in the specific context of one-shot games (i.e. selfishness). Direct evidence of such spillovers comes from a study in which subjects play a series of either long or short repeated PDs, and then a battery of one-shot anonymous games (Peysakhovich and Rand in press). Subjects randomized into the long PD condition are dramatically more pro-social in the subsequent one-shot games compared to subjects randomized into the short PD condition.

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 strategy of tit for tat, reciprocal interactions should lead to intuitions that favor cooperation at the outset of a relationship, and cooperation in response to a cooperative partner. But 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 rather than intuitive cooperation.

Support for this prediction comes from experiments using 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, while reflection leads to increased acceptance (Gospic et al. 2011; Grimm and Mengel 2011; Sanfey et al. 2003; Sutter et al. 2003) (although evidence from transcranial magnetic stimulation experiments suggests that deliberative processes also play some role in rejections; Knoch et al. 2006; Wout et al. 2005). Thus intuition again favors reciprocation (in this case, paying a cost to retaliate against selfishness). As with cooperation in one-shot interactions, rejecting unfair offers in the UG is not payoff maximizing in the one-shot games studied in the lab, but is adaptive in the context of reciprocal interactions (Nowak et al. 2000).

This evidence of intuitive reciprocation supports the argument that strategies selected in the context of repeated games spill over into one-shot anonymous interactions. But are these intuitions the result of genetic hardcoding, 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 (Hauge et al. 2009; Rand et al. 2013b; Tinghög et al. 2013), 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 where most others are trustworthy and cooperative themselves (Rand et al. 2012; Rand and Kraft-Todd 2014). If you grow up in a non-cooperative equilibrium, where cooperation is not payoff maximizing, you internalize defection as your default. Prior experience with
behavioral experiments also moderates the role of intuition in cooperation. Individual-differences studies show that intuitive responses are more cooperative among naïve subjects, but that intuition does not promote cooperation among experienced subjects (Rand et al. 2012; Rand et al. 2014; Rand and Kraft-Todd 2014). At the study level, the effect of an increasingly experienced subject pool was explored by analyzing a series of experiments conducted over two years using the online labor market Amazon Mechanical Turk (Rand et al. 2014). During that period, behavioral experiments became dramatically more common on Mechanical Turk, resulting in a subject pool that is highly experienced with study participation. As predicted by the Social Heuristics Hypothesis, decisions made under time pressure became steadily less cooperative (as intuitions were eroded), while reflective responses remained constant. These findings suggest that intuitions are malleable rather than hardcoded. Thus we find support for the Social Heuristics Hypothesis and for the importance of learning and culture in human cooperation.

Conclusion

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 where 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, multi-level selection, and kin selection. We also highlight areas where theory and experiments diverge and where more empirical and theoretical work is needed.

We also consider cooperation in one-shot anonymous settings where 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 toward reciprocal 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 (1) some level of one-shot cooperation was specifically favored by selection, or that (2) genetic evolution played an important role in the evolution of human cooperation (in support of the latter point, for example, babies seem to have reciprocal preferences [Hamlin et al. 2007; Hamlin et al. 2011] and young children are often willing to help others [Warneken and Tomasello 2006, 2009; Warneken et al. 2007]). Conducting experiments to further distinguish between these hypotheses for the origins of human cooperation in one-shot interactions is a fundamental challenge for the field. Critically, all of the perspectives on the evolution of human cooperation outlined in this chapter
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, particularly in the economics literature. Behavior in the laboratory cannot be explained without considering the environment in which that behavior evolved.

Our conclusion regarding the important of reciprocity for human cooperation complements the recently proposed 'hexagon of cooperation' (Messner, Guarín, and Haun in this volume). Here we have discussed the five mechanisms for the evolution of cooperation, interaction structures that can allow selection to favor cooperation when added to the PD. Rather than examining mechanisms per se, the hexagon of cooperation explores factors that are important for the successful functioning of mechanisms involving reciprocity. Some of these factors may be particularly important for direct reciprocity, such as trust. Others may be particularly important for indirect reciprocity, such as communication. Thus the hexagon of cooperation can provide a useful set of guidelines for those trying to use reciprocity to promote human cooperation.

Finally, we consider the implications for each mechanism of the modern world’s ever-increasing scale and complexity of interaction. Spatial selection and multi-level selection rely on segregation: clusters of cooperators supporting each other, or groups of cooperators outcompeting groups of defectors. The global interconnections and interdependencies of the 21st century are therefore challenging for these mechanisms, making it harder for them to promote cooperation. Yet the opposite is true of mechanisms based on reciprocity (and, in particular, indirect reciprocity). Greater interdependence increases the power of reciprocity, both by increasing information flow regarding others’ past behavior and by increasing the value of maintaining good relationships in the future. Thus, consistent with the hexagon of cooperation (Messner et al. in this volume), we may see an ever greater emphasis on reciprocity as our world becomes increasingly complex and interdependent. Although many may believe that selfishness is the essence of the human condition, this is not so: in a world where we depend on each other ever more, cooperation is clearly the winning strategy.

Note
1 Funding from the John Templeton Foundation is gratefully acknowledged.

References


