Evolving the Psychological Mechanisms for Cooperation

Jeffrey R. Stevens  
*University of Nebraska-Lincoln*, jstevens5@unl.edu

Fiery A. Cushman  
*Harvard University*, fiery_cushman@brown.edu

Marc D. Hauser  
*Harvard University*

---

Follow this and additional works at: [https://digitalcommons.unl.edu/psychfacpub](https://digitalcommons.unl.edu/psychfacpub)

Part of the Psychiatry and Psychology Commons

Stevens, Jeffrey R.; Cushman, Fiery A.; and Hauser, Marc D., "Evolving the Psychological Mechanisms for Cooperation" (2005). *Faculty Publications, Department of Psychology*. 540.  
[https://digitalcommons.unl.edu/psychfacpub/540](https://digitalcommons.unl.edu/psychfacpub/540)
Evolving the Psychological Mechanisms for Cooperation

Jeffrey R. Stevens, Fiery A. Cushman, and Marc D. Hauser

Primate Cognitive Neuroscience Laboratory, Department of Psychology, Harvard University, Cambridge, Massachusetts 02138

Abstract
Cooperation is common across nonhuman animal taxa, from the hunting of large game in lions to the harvesting of building materials in ants. Theorists have proposed a number of models to explain the evolution of cooperative behavior. These ultimate explanations, however, rarely consider the proximate constraints on the implementation of cooperative behavior. Here we review several types of cooperation and propose a suite of cognitive abilities required for each type to evolve. We propose that several types of cooperation, though theoretically possible and functionally adaptive, have not evolved in some animal species because of cognitive constraints. We argue, therefore, that future modeling efforts and experimental investigations into the adaptive function of cooperation in animals must be grounded in a realistic assessment of the psychological ingredients required for cooperation. Such an approach can account for the puzzling distribution of cooperative behaviors across taxa, especially the seemingly unique occurrence of cooperation observed in our own species.

1. Introduction

Vampire bats regurgitate blood to others despite the possibility of dying if three days elapse without consuming blood. Ground squirrels give alarm calls even though they alert predators to their own presence. Cleaner fish enter the mouths of their hosts to remove parasites even at risk of being eaten. Florida scrub jays often stay at home with their parents, foregoing the benefits of personal reproduction to help rear their younger siblings. These cases of cooperation have generated a substantial amount of theoretical and empirical interest over the past several decades, primarily focusing on adaptive accounts of cooperative behaviors. This adaptive perspective has been fruitful; indeed, the crowning glory of the sociobiological revolution beginning in the 1960s has been the overwhelming empirical support for its theoretical predictions targeted at adaptive accounts of social behavior (Alcock 2001, Hamilton 1964, Trivers 2002, Williams 1966, Wilson 1975). The adaptive view, however, fails to
fully account for the empirical data on cooperative behavior. This weakness, in our opinion, results from a strictly ultimate perspective that ignores proximate mechanisms of cooperation. Here, we argue that evolutionary puzzles concerning the phylogenetic distribution of cooperative behaviors can be resolved by unraveling the psychological machinery upon which they depend.

We first discuss the adaptive challenges of cooperation and briefly review potential solutions. We then describe the empirical evidence for the different solutions, demonstrating that some types of cooperation occur much less frequently in nonhuman animals (hereafter animals) than theory predicts. Finally, we introduce the idea of cognitive constraints on cooperation to explain the limited taxonomic scope of certain kinds of cooperation. We conclude by discussing how a psychologically informed approach to cooperation opens a new set of questions, guides how we design our experiments, and helps resolve apparently contradictory findings concerning the uniqueness of human cooperation.

2. The Problem Of Cooperation

We define cooperation as any behavior that provides a benefit to an individual other than the cooperator (where benefit is defined as an increase in reproductive success). Given the generally accepted Darwinian assumption that behavior evolves via natural selection, cooperation poses an apparent problem: What selective pressure favors individuals who provide benefits to other individuals? A number of models address this problem, of which we review four: mutualism, kin selection, reciprocity, and sanctioning (for more complete treatment, see Dugatkin 1997).

2.1 Mutualism

The simplest explanation for cooperative behavior is that it provides direct benefits to the cooperator, in addition to other individuals. This model of cooperative behavior is termed mutualism (Brown 1983, West Eberhard 1975). Any individual that defects (i.e., does not cooperate) in mutualistic situations will, by definition, do worse than a cooperator; therefore, in the absence of a temptation to defect, cooperation provides the best option. Importantly, mutualism does not depend on the identity of your partner and, therefore, can occur between any members of the same species and even members of different species (Boucher 1985, Herre et al. 1999).

2.2. Kin Selection

Cooperation that does not yield the direct benefits of mutualism poses an even deeper evolutionary paradox: altruism. Why would an individual help others, especially at a cost to itself? Kin selection provided the first clear theoretical solution to the paradox of altruism. Although introduced by Darwin (1859), Hamilton (1964) first mathematically formalized kin selection as a mechanism to main-
tain cooperation among genetic relatives. He suggested that individuals may bias cooperation toward their genetic relatives because it helps propagate their own genes. What looks altruistic from an individual’s perspective actually serves self-interest from the gene’s view (Dawkins 1976). Individuals share a certain proportion of their genes ($r$ - the coefficient of relatedness) with relatives due to common descent. If the benefits to kin discounted by this coefficient of relatedness outweigh the costs of helping, altruism toward kin can evolve.

2.3. Reciprocity

Reciprocity, in which individuals pay a short-term cost of cooperation for the future benefit of a social partner’s reciprocated cooperation, has probably been the most celebrated type of cooperation. Reciprocity aims to explain cooperative behavior in a unique type of social interaction termed the prisoner’s dilemma (Flood 1958, Rapoport & Chammah 1965). The key aspects of the prisoner’s dilemma are (a) cooperation maximizes the total payoff to everyone involved in the interaction (mutual cooperation provides more benefits than mutual defection); however, (b) any individual will receive a higher personal payoff by defecting, so a sizable temptation to cheat exists (Figure 1a). Pursuing unilateral cooperation in this game is not an evolutionarily stable strategy (Maynard Smith 1982, Maynard Smith & Price 1973).

Trivers (1971) suggested that reversing roles as donor and recipient of altruism may reduce the temptation to defect because individuals are investing in future cooperation. Reciprocity can stabilize cooperation if the following conditions are met: (a) the benefits to the recipient outweigh the costs to the donor, (b) individuals interact repeatedly, and (c) individuals recognize partners so they can detect cheaters. When the fitness payoffs sum over a series of interactions with the same partner, reciprocal strategists can reap the benefits of mutual cooperation (Figure 1b). The reciprocal strategy tit-for-tat (TFT), in which a player starts out cooperating and copies its opponent’s behavior in previous interactions, can successfully invade and dominate simulated populations of social partners engaging in prisoner’s dilemma games, winning out over many alternative behavioral strategies (Axelrod 1984, Axelrod & Hamilton 1981). If the probability of interacting again exceeds a critical level, a reciprocal strategy can maintain cooperation.

2.4. Sanctioning

Punishing defection can impose enough costs to offset the temptation to cheat and, like reciprocity, can elicit future cooperation (Boyd & Richerson 1992, Clutton-Brock & Parker 1995a). Punishment involves energetic costs and, when accomplished by aggression, also involves the cost of risked injury. Punishment can only be an adaptive behavior at the individual level of selection, therefore, when it successfully elicits cooperative behavior directed strictly at the punisher (Gardner & West 2004). In this manner punishment resembles reciprocity, which should occur when it elicits cooperative behavior directed strictly at the reciprocator.
Another form of sanctioning is harassment. Whereas punishment penalizes past behavior with the hope of future reward, harassment penalizes present behavior with the hope of present reward. For instance, if a defector has consumed 100% of a food resource, a punisher could punish the defector in the hope of obtaining 50% of the food resource in some future interaction; however, if a defector is in the process of consuming a food resource, a harasser could impose costs on the defector aimed at obtaining 50% of the food resource immediately. By harassing during defection, individuals may induce cooperation, thereby providing an immediate individual benefit for the harasser, rather than the delayed benefit of future cooperation (Stevens & Stephens 2002).
3. Evidence for Cooperation

Despite the abundance of theoretical interest in the different models of cooperation, over 30 years of research on a wide variety of species, under lab and field conditions, reveals that some occur much more frequently in animals than others. In particular, mutualism and kin-biased cooperation account for the vast majority of instances of cooperation. In stark contrast, only a handful of species have demonstrated reciprocity, and even within these species reciprocity occurs infrequently (Hammerstein 2003, Noë 1990, Stevens & Hauser 2004). Sanctioning is also inconsistently distributed across taxa; punishment appears to be less common than mutualism or kin selection, but harassment may occur frequently, perhaps especially among primates (Stevens & Gilby 2004). Here we describe only a fraction of the empirical evidence for the four models of cooperation. The goal is to look at a few selective cases and then attempt to explain the patterns in light of our argument for cognitive constraints.

3.1. Mutualism

Mutualism may be the most common form of cooperation; it occurs frequently across a wide variety of taxa. Cooperative hunting provides a prime example described in numerous species (Dugatkin 1997, Krause & Ruxton 2002). Cooperative hunting provides mutualistic benefits only when the per capita intake rate increases with group size. Therefore, a pair of hunters would have to capture more than twice as many prey items as a solitary hunter. For cooperative hunting to benefit hunters, the success rate of solitary hunters should be fairly low, making cooperation particularly successful (Packer & Ruttan 1988). In addition to increasing the probability of capture, cooperative hunting may also reduce the individual costs of hunting (Creel 1997). Combined, these two mechanisms can lead to direct, immediate, and simultaneous fitness benefits for cooperative hunters, particularly when hunting small or difficult prey.

Another example of mutualism is food recruitment in ravens (Corvus corax). Heinrich and colleagues (Heinrich 1989, Heinrich & Marzluff 1991) observed that when ravens discovered animal carcasses, they often gave food calls and returned to communal roosts to recruit others. After ruling out reciprocity and kin-biased cooperation (Heinrich 1988, Parker et al. 1994), Heinrich proposed that the callers recruited others to prevent territory owners from monopolizing the food. Therefore, though it appears altruistic, recruiting actually increased access to an otherwise unavailable food source.

Cooperative breeders may also accrue mutualistic benefits (Woolfenden & Fitzpatrick 1978). Individuals help raise their younger siblings because larger group sizes can yield benefits such as larger territories. These territories often have better access to food, reduced predation risk, and space to establish their own territories (Clutton-Brock 2002, Cockburn 1998, Stacey & Koenig 1990). Some species even adopt or kidnap unrelated offspring, presumably to reap the benefits of increasing their group size (Connor & Curry 1995, Heinsohn 1991, Zahavi 1990).
3.2. Kin Selection

Kin-biased cooperation commonly occurs in species ranging from arthropods to apes (Bourke 1997, Chapais & Berman 2004, Silk 2002). Darwin’s original paradox of altruism was motivated by the kin-biased cooperation that occurs in eusocial insects such as ants, bees, and wasps (Bourke 1997). In eusocial species, individuals cooperate to care for the young, divide reproductive opportunities such that only a few individuals reproduce, and have overlapping generations (Hölldobler & Wilson 1990). Importantly, many individuals forego reproduction completely to aid in the care of their siblings. Some species even express haplodiploid genotypes with haploid males and diploid females. This results in sisters being highly related to each other \((r = 3/4)\) but less closely related to their brothers \((r = 1/4)\) and even their own daughters and sons \((r = 1/2)\); therefore, females should bias more cooperation toward other sisters. In a now classic study, Trivers & Hare (1976) used this unique system to test kin selection in ants. They found that females matched their altruistic allocations to the degree of genetic relatedness: Investment in other females exceeds investment in males by three orders of magnitude (for an alternative perspective, see Reeve 1993).

Kin selection may be particularly powerful in cooperatively breeding species, although it by no means accounts for all cooperative behaviors in these systems (Clutton-Brock 2002, Cockburn 1998, Griffin & West 2002). Kin selection theory predicts that individuals should preferentially help relatives and should help in proportion to their coefficient of relatedness. Reviews of the literature support this prediction, demonstrating that the probability and amount of help correlate with the coefficient of relatedness (Emlen 1997, Griffin & West 2003). Griffin & West (2003) tested even stronger predictions in a meta-analysis by demonstrating that the effect of kin selection (i.e., the correlation between helping and relatedness) correlates highly with the benefit of helping. This meta-analysis extended the findings of Trivers & Hare (1976) to other species, suggesting that helping scales with not only the coefficient of relatedness \((r)\) but also the benefit of helping.

3.3. Reciprocity

Many authors have reported reciprocity—the alternation of receiving costs and benefits—in numerous contexts including food sharing (de Waal & Berger 2000, Hauser et al. 2003, Watts & Mitani 2002), grooming (Barrett & Henzi 2001, Rowell et al. 1991), predator inspection (Dugatkin 1988, Milinski 1987), and coalitions associated with mating opportunities (Packer 1977). Unfortunately, most examples of reciprocity suffer from one of two problems: (a) they have not been replicated, and (b) alternative explanations, such as kin selection and mutualism, can account for the observed reciprocal pattern.
One of the first reported instances of reciprocity occurred in olive baboons (*Papio anubis*) (Packer 1977). Males formed coalitions in order to drive off rival males and gain access to reproductively active females. Packer’s analyses suggested that males took turns reaping the benefits, implying evidence of reciprocity. A subsequent study of a different population of baboons did not find the same reciprocal patterns (Bercovitch 1988), and a study of a closely related species proposes an alternative explanation. Noë (1990) suggested that male savannah baboons (*Papio cynocephalus*) implemented multiplayer market strategies during coalition formation. Coalition partners did not face a prisoner’s dilemma; rather they faced a veto game in which the coalition initiator forces other males to “bid” on joining the coalition. If the potential helpers did not join, they would not receive mating opportunities. Therefore, the coalition initiator can choose the partner which will take the lowest share of mating opportunities.

Probably the best-known putative example of reciprocity is Wilkinson’s (1984) study of blood sharing in vampire bats (*Desmodus rotundus*). In this study, vampire bats regurgitated blood to individuals that failed to obtain blood meals on their own. The bats shared mainly with individuals with whom they associated, leading Wilkinson to conclude that the bats reciprocated. Although commonly cited as evidence for reciprocity, only 5 of the 98 instances of sharing between individuals of known genetic relatedness involved individuals less related than grandparent to grandchild (*r* < 0.25); most occurred between mother and offspring. Therefore, direct fitness benefits and kin selection can account for the vast majority of sharing, leaving instances of potential reciprocal sharing quite rare and possibly attributable to recognition errors.

Another controversial example of reciprocity is predator inspection in stickleback fish (*Gasterosteus aculeatus*) and guppies (*Poecilia reticulata*). Milinksii (1987) and Dugatkin (1988) found that when inspecting a predator fish, individuals copied their partners’ approach or retreat (the “partners” were actually mirrors that mimicked a fish either swimming with or away from the subject). These experiments elicited a flood of criticisms. In particular, the fitness payoffs of cooperation and defection were unclear (Connor 1996, Lazarus & Metcalfe 1990), and rather than reciprocating, the target fish may simply have preferred to stay in groups to reduce predation risk (Masters & Waite 1990, Stephens et al. 1997). Minimally, we consider the stickleback and guppy work on reciprocity as unresolved.

Finally, we describe several cases that appear to offer viable evidence of reciprocity. Hauser et al. (2003) conducted a series of experiments in which cotton-top tamarins (*Saguinus oedipus*) could altruistically pull a tool to give food to an unrelated recipient without getting any food for itself. Subjects alternated which partner had the opportunity to pull with a short time interval between trials. Tamarins pulled the tool most often for partners that always pulled and infrequently for partners that never pulled. The tamarins, however, cooperated less than 50% of the time, and as each game progressed, the amount of food given decreased. Tamarins, therefore, maintained a moderate level of cooperation when receiving food closely followed giving food.
Some researchers have proposed reciprocal egg swapping in simultaneously hermaphroditic fish (chalk bass—Serranus tortugarum) and polychaete worms (Ophryotrocha gracilis) as examples of reciprocity (Fischer 1988, Sella et al. 1997). These species produce both male and female gametes and, therefore, can both give and receive fertilizations. Because eggs cost more to produce, a defector could fertilize a cooperator’s eggs but offer none of its own, thereby avoiding costly egg production. Both the fish and worms repeatedly alternate depositing packets of eggs and fertilizing their opponents’ packets of eggs. Depositing eggs depends on the partner’s behavior because the interaction stops when the partner fails to deposit eggs. Reciprocal allogrooming in impala (Aepyceros melampus) follows a similar pattern: Individuals groom one another for short bouts and then receive grooming from their partner, repeatedly alternating who grooms (Hart & Hart 1992). In all of these examples, the small time delay between paying the cost of cooperation and receiving the benefit minimizes the chance of defection. In Section 4.2, we further discuss the significance of time in the evaluation of reciprocal possibilities.

3.4. Sanctioning

Despite the theoretical interest, punishment is not well documented in animals. Clutton-Brock & Parker (1995a) offered several examples of punishment enforcing cooperation in animals, but we would characterize these examples as harassment because the punisher usually receives immediate benefits from punishing. For instance, queen paper wasps (Polistes fuscatus) attack lazy workers; when Reeve & Gamboa (1987) removed the queen from the colony, the workers stopped working. Although reported as punishment, one should categorize this as harassment because the queen’s aggression immediately increased the activity of the workers.

Animals often impose costs on others to influence their current behavior (Clutton-Brock & Parker 1995b, Stevens 2004, Wrangham 1975). For instance, Gilby (2004) studied the food-sharing patterns of wild chimpanzees (Pan troglodytes). After capturing prey, the chimpanzees frequently allowed other individuals to consume part of the meat. Gilby showed that harassment accounted for the pattern of food sharing, because harassment was costly for the food owner (food intake rate decreased as the number of beggars increased), owners shared more often when beggars harassed frequently and intensely, and when sharing occurred harassment levels decreased. Controlled experiments corroborated these findings with captive chimpanzees and extended them to other species such as squirrel monkeys (Saimiri boliviensis), a species that rarely cooperates (Stevens 2004).

Harassment may, of course, influence future as well as current cooperation, suggesting that it may lead to punishment strategies. For example, when rhesus monkeys (Macaca mulatta) discovered food, those individuals who announced their discovery by vocalizing faced fewer attacks than those individuals who made their discoveries in silence, apparently withholding information (Hauser 1992, Hauser & Marler 1993). This sanctioning imposed costs on the
discoverer not only in terms of potential for injury and wasted energy but also by reduced food intake relative to vocal discoverers. This sanctioning yielded an immediate benefit of accessing food, clearly qualifying it as a case of harassment. In addition, these data suggest that the sanctioning may have had a punishing effect on silence, eliciting future food calls. That is, sanctioning behavior resulted in both immediate and future benefits. To the authors’ knowledge, no clear evidence in animals demonstrates that punishment influences future cooperation in the absence of harassment.

In summary, although ample evidence of cooperation exists in the animal kingdom, reciprocity and punishment rarely occur in animals, especially when contrasted with humans (Fehr & Gächter 2002, Ostrom & Walker 2003). We next turn to an explanation for this taxonomic distribution.

4. Cognitive Constraints on Cooperation

Mutualism and kin selection are both theoretically well understood and empirically well documented (Dugatkin 1997). By contrast, reciprocity and punishment, although theoretically feasible, do not frequently occur in animals (although harassment may be more widespread). Therefore, despite models purporting the evolutionary stability of all of these types of cooperation, some types occur much more frequently than others. Unfortunately, a strictly adaptive perspective has limited power to explain the frequency of mutualism and kin-biased cooperation, and the rarity of reciprocity and punishment. A proximate perspective that keeps its eye on the ultimate problem can, however, reveal how psychological constraints limit or facilitate particular forms of cooperation.

The proximate approach emphasizes critical aspects of reciprocity and punishment that differ markedly from mutualism and kin-biased cooperation. In both reciprocity and punishment, the fitness benefits associated with cooperation depend on the partner’s behavior: Cooperation should only occur when the partner responds by reciprocating or punishing. When this contingent response occurs in the future, the temporal delay introduces cognitive challenges that may constrain the emergence and stability of cooperation (Stevens & Gilby 2004). Animals can easily implement strategies that yield immediate benefits, such as mutualism and harassment, because individuals do not have to track benefits over time. With a time delay between cooperating and receiving return benefits, however, individuals must invest in an uncertain future. Delayed benefits impede learning the consequences of cooperation, require more memory capacity for previous interactions, and trade off short-term fitness gains for long-term gains. Here we provide a sketch of our proximate perspective on cooperation, highlighting several key cognitive constraints as a way to distinguish mutualism, kin-biased cooperation, reciprocity, and sanctioning.

Because no temptation to cheat exists in mutualistic interactions, individuals should always cooperate. As a result, mutualism requires no special cognitive abilities above and beyond the challenges inherent in the cooperative behavior itself. Although Dugatkin & Alfieri (2002) contend that animals must recognize whether they are in a mutualistic situation, we argue that recognition is neither necessary for nor specific to mutualism. Kin-biased cooperation, on the other hand, does require additional cognitive capacities. At a minimum, it requires the capacity to direct cooperative actions to related individuals (Dugatkin & Alfieri 2002). Indeed, Hamilton’s formulation of kin selection spawned a critical series of empirical studies showing that numerous species had the ability to make certain kin discriminations (Fletcher & Michener 1987, Hepper 1991). These studies were critical in the general acceptance of kin selection.

Mechanisms of kin recognition include recognition alleles, phenotype matching, and spatial and familiarity cues (Hepper 1991, Sherman et al. 1997, Wilson 1987). The recognition allele hypothesis predicts that individuals can compare a particular phenotypic cue (auditory, olfactory, visual, etc.) to an innately specified template (e.g., the “green beard effect” – Dawkins 1976, Hamilton 1964). Such a model requires few cognitive skills other than discriminating the cue associated with relatedness. Keller & Ross (1998) suggested that fire ants (Solenopsis invicta) may use recognition alleles to selectively kill queens that do not share their genotype. Phenotypic matching occurs when an individual compares a conspecific’s phenotypic cues to a learned template. This requires specialized perceptual and computational systems that detect cues at an early stage to form a template, then test cues against the template to discriminate kin (Hauber & Sherman 2001). Peacocks (Pavo cristatus), golden hamsters (Mesocricetus auratus), and a number of other species have demonstrated phenotypic matching (Mateo & Johnston 2000, Petrie et al. 1999). Finally, a common alternative is to use a simple set of rules such as spatial and familiarity cues to discriminate kin. Often, animals may use rules such as “be nice to individuals near your home” or “help those that you grew up with” to direct the benefits of cooperation toward kin. These mechanisms occur regularly in a variety of animal taxa.

4.2. Cognitive Constraints on Reciprocity and Punishment

Trivers’ (1971) classic formulation of reciprocity had three requirements for evolutionary stability: (a) the reciprocated benefit must outweigh the immediate cost, (b) individuals must interact repeatedly, and (c) individuals must recognize each other. We contend that these requirements do not capture the cognitive sophistication required for utilizing reciprocal strategies. In particular, the delay between the cost of a cooperative act and the benefit of reciprocated cooperation introduces a number of cognitive challenges. Like reciprocity, punishment can involve a delay between a costly act and a beneficial payoff, and in these cases it faces similar constraints. For this reason, we consider the constraints on reciprocity and punishment together.
4.2.1. **Individual Recognition**  
Trivers does propose individual recognition as a cognitive ability needed to avoid cheaters and stabilize cooperation via reciprocity (see also Dugatkin & Alfieri 2002). Punishment also requires individual recognition to ensure that individuals only cooperate with punishers, thereby preventing punishment from benefiting nonpunishers (Gardner & West 2004). Therefore, the delayed, contingent response required for both reciprocity and punishment necessitates that individuals can distinguish different partners. Numerous species across the animal kingdom possess the ability to recognize individuals, however, so the necessity for a mechanism of individual recognition cannot explain the paucity of cooperative behavior across most nonhuman taxa.

4.2.2. **Temporal Discounting**  
Temporal discounting is the devaluing of future rewards, which often results in a preference for smaller, immediate rewards over larger, delayed rewards. Many psychologists who study discounting consider the prisoner’s dilemma to be analogous to the discounting problem (Green et al. 1995, Rachlin 2000). Individuals must choose between the immediate reward of defecting and the long-term reward of cooperating. Indeed, a number of researchers have predicted that temporal discounting can reduce the value of reciprocated benefits (Frank 1988, May 1981, Trivers 1971). Experimental data on variation in human discounting and cooperation validate the view that a preference for immediate rewards may inhibit reciprocity. Discounting correlates with cooperation such that individuals who prefer immediacy cooperate less frequently (Harris & Madden 2002). In parallel, blue jays (*Cyanocitta cristata*) show stable cooperation in the prisoner’s dilemma only following a reduction in their preference for immediacy resulting from accumulating payoffs over several trials (Stephens et al. 2002). Therefore, if animals highly discount future rewards (Figure 2; Mazur 1987, Richards et al. 1997, Stevens et al. 2005), the immediate benefits of defections may outweigh the future reciprocated benefits.

4.2.3. **Memory**  
Limitations in memory decay, interference, and capacity can also constrain the frequency of reciprocity and punishment. Models of forgetting predict exponential or power functions (Sikstrom 2002, White 2001, Wixted 2004), because memories decay rapidly over time (Figure 3). Therefore, longer time intervals between cooperative acts may make reciprocity and punishment more difficult. Even with short time delays between cooperative interactions and few distractions, every potential new partner increases the computational load of tracking debts owed, favors given, and costs imposed. Keeping score of reciprocal obligations and punishment with multiple individuals may place a computationally intensive burden on memory systems. Although few studies examine learning and memory constraints in animal cooperation, human studies suggest that these constraints can pose challenges for maintaining stable cooperative relationships (Milinski & Wedekind 1998).

Existing evidence demonstrates the importance of the time delay between paying the costs of cooperation and receiving the benefits. In the previously described examples of food sharing in tamarins, reciprocal egg swapping
in fish and worms, and reciprocal allogrooming in impala, the time delay is minimal. This greatly reduces the cognitive demands for reciprocity: Individuals no longer need to recognize each other because they are always in proximity during the interactions, the benefits accrue immediately, and thus avoid discounting, and memory is not needed because individuals can leave when their partner cheats.

4.3. Cognitive Constraints on Harassment

Like mutualism and kin-biased cooperation, harassment does not suffer from the same cognitive limitations as reciprocity and punishment, primarily because of the brief time delay between imposed costs and potential cooperation. Harassment does not require individual recognition because individu-
Individuals need not interact repeatedly; harassment can elicit cooperation between perfect strangers. Temporal discounting plays a minor role because harassers receive immediate benefits to offset their own costs of harassing. In addition, eliminating the time delay removes limitations associated with memory. Although harassment is less well studied than the other forms of cooperation, we predict that it occurs frequently in the animal kingdom given its weak demands on psychological capacities and its utility in reaping benefits at a small cost.

5. Conclusions

Several different models have solved the adaptive paradox posed by cooperative behavior; here, we have reviewed mutualism, kin selection, reciprocity, and sanctioning. Though empirical evidence for mutualism and kin-biased co-

**Figure 3.** Species comparison of memory. Comparisons of memory across species is probably even more problematic than discounting, again because of methodological differences. Stumptail macaques (*Macaca arctoides*) and pigeons show fairly steep decreases in memory retention in binary delayed matching-to-sample tasks (Jarrad & Moise 1970, Wixted & Ebbesen 1991). Although not a perfectly analogous comparison, face recognition tasks in humans show a much longer retention interval (Wixted & Ebbesen 1991).
operation is widespread, reciprocity occurs relatively infrequently among non-human animals. Many of the current studies on sanctioning in animals are better explained by harassment models than by punishment models. We propose that a unique feature of reciprocity and punishment can explain this mismatch between the models and the data: Individuals must balance the costs and benefits of repeated interactions over periods of delay. Balancing costs and benefits over time poses several cognitive challenges, and therefore the emergence of reciprocity and punishment faces nontrivial psychological constraints. Here we focus on the possible role of individual recognition, temporal discounting, and memory as specific constraints on reciprocity and punishment. This is by no means an exhaustive list of abilities required to implement these strategies. Instead, this approach raises important questions about the nature of cooperation. What other abilities does cooperation require? Does incorporating these constraints lead to more predictive power in models of cooperation? How do psychological and evolutionary prerequisites interact to allow the implementation of reciprocal and punishment strategies? How do these and other factors influence not only animal cooperation but also the ontogeny of cooperation in our own species?

Knowing that a variety of psychological mechanisms facilitate cooperation allows us to design more appropriate experiments. Investigations of reciprocity and punishment must first evaluate the limitations of animals in the areas we have described, as well as others. What are the recognition abilities, discounting rates, memory features, numerical discrimination abilities, and learning rates of the species being investigated? For example, given evidence of limits on number discrimination (Dehaene 1997, Gallistel 1990, Hauser 2000), it makes little sense to set up payoff matrices that entail nondiscriminable alternatives. When testing reciprocity and punishment, researchers must first consider the cognitive constraints operating in the target animal and, based on this analysis, design an appropriate experiment.

We emphasize that the cognitive constraints discussed in this paper pose a challenging hurdle to the evolution of reciprocity and punishment but not an insurmountable barrier. Instances abound of extraordinary cognitive adaptations narrowly tailored to specific behavioral routines, overcoming the more general cognitive limitations. For instance, Clark’s nutcrackers (*Nucifraga columbiana*) stash several thousand seeds each fall, foregoing the opportunity of immediate consumption in favor of future benefits. This behavior is grossly inconsistent with the typical rate of temporal discounting in nonhuman animals, implicating a specific cognitive adaptation bypassing a more general constraint. The nutcracker also depends on extraordinary spatial memory, which greatly facilitates its capacity to retrieve stashed food from thousands of hiding locations several months later (Balda & Kamil 1992, Kamil et al. 1994). These impressive cognitive adaptations enable a single, specific behavioral routine. By analogy, we should expect to find reciprocity and punishment in instances where adaptation has overcome the initial cognitive constraints—where narrowly tailored cognitive mechanisms have evolved to support specific behavioral routines.
An exception to this pattern, gratifyingly, seems to prove the rule. In humans, reciprocity and punishment commonly occur across a broad array of social interactions (Camerer 2003, Fehr & Gächter 2002, Fehr et al. 2002, Gurven 2004, Ostrom & Walker 2003, Ostrom et al. 1992). Far from being narrowly tailored to specific behavioral routines, reciprocity and punishment are broad and flexible strategies that can be applied to novel circumstances. Traditional, ultimate models fail to explain the ubiquity of cooperation among humans, where factors like population size, migration rates, frequency of interaction, and the cost-benefit structure of social interaction play the key roles in constraining or enabling reciprocity and punishment. By these measures, nothing about humans is very unique. Factoring in the role of cognitive constraints, however, helps explain the uniqueness of human cooperative behavior. In important ways, human cognition differs from nonhuman cognition and may pose fewer constraints on the emergence of reciprocity and sanctioning. For instance, rats, pigeons, and even nonhuman primates devalue rewards postponed by just a few seconds; to see this kind of discounting function in humans requires extending choices over months rather than seconds (Figure 2; Mazur 1987, Rachlin et al. 1991, Richards et al. 1997, Stevens et al. 2005). As a result, humans do not need to evolve narrowly tailored exceptions to their general rate of temporal discounting because it does not impose a severe constraint on reciprocity or punishment. Other features of human cognition that may enable reciprocity and punishment include face recognition and episodic memory, allowing for specific social interactions to be recalled; language, allowing for the negotiation of threats and promises and for facilitated bookkeeping by tagging cooperators and cheaters with symbols or labels; and theory of mind, allowing for inferences of intent and motivation in social exchange. Although some of these cognitive capacities are shared in part or in whole with nonhuman animals, others appear unique to humans. In his original formulation of reciprocity, Trivers (1971) emphasized the myriad cognitive abilities that humans may use to implement reciprocal strategies. We, however, disagree with Trivers’ suggestion that these abilities evolved after reciprocity as regulating mechanisms. Instead, we concur with Darwin (1872) and Williams (1966) in that reciprocity requires the existence of these faculties before it can evolve.

Why have no other species evolved these mechanisms to allow reciprocity? The effort poured into the theoretical analyses of reciprocity may not reflect its frequency in the wild. Animals may not face altruistic situations in which they interact repeatedly with nonkin. Many instances of cooperation that appear altruistic may, instead, provide immediate mutualistic benefits such as the raven food-calling example. Similarly, individuals interact with genetic relatives so often that kinship may drive the majority of their social situations (e.g., vampire bat blood sharing). With few opportunities for reciprocity to provide benefits, selection may have been too weak to overcome cognitive constraints.

Mutualism, kin selection, reciprocity, and sanctioning elegantly explain how we can reconcile cooperative behavior against the Darwinian maxim that selection favors behavior that maximizes personal gains in terms of survival and reproduction. Developed in response to an adaptive paradox, how-
ever, these models have long neglected the role of mechanistic constraints. Integrating animal psychology into current models of cooperative behavior can help explain the curious taxonomic distribution of reciprocity and punishment, which are rare among nonhuman animals but ubiquitous among humans. We have touched upon a few of the possible cognitive constraints on cooperation, and eagerly anticipate future research to expand and clarify the role of others.

Acknowledgments — We gratefully acknowledge funding from the National Institutes of Health (NRSA for J.R.S.) and the National Science Foundation (ROLE grant for M.D.H.). We appreciate comments on the manuscript from Mike Ryan, David Stephens, and Robert Trivers.

Literature Cited

Darwin C. 1872. The Descent of Man and Selection in Relation to Sex. London: Murray


