Why humans might help strangers

Nichola J. Raihani[1]* and Redouan Bshary[2]

[1] Department of Genetics Evolution and Environment, University College London, London, UK
[2] Institut de Biologie, Université de Neuchâtel, Neuchatel, Switzerland

Humans regularly help strangers, even when interactions are apparently unobserved and unlikely to be repeated. Such situations have been simulated in the laboratory using anonymous one-shot games (e.g., prisoner's dilemma) where the payoff matrices used make helping biologically altruistic. As in real-life, participants often cooperate in the lab in these one-shot games with non-relatives, despite that fact that helping is under negative selection under these circumstances. Two broad explanations for such behavior prevail. The "big mistake" or "mismatch" theorists argue that behavior is constrained by psychological mechanisms that evolved predominantly in the context of repeated interactions with known individuals. In contrast, the cultural group selection theorists posit that humans have been selected to cooperate in anonymous one-shot interactions due to strong between-group competition, which creates interdependence among in-group members. We present these two hypotheses before discussing alternative routes by which humans could increase their direct fitness by cooperating with strangers under natural conditions. In doing so, we explain why the standard lab games do not capture real-life in various important aspects. First, asymmetries in the cost of perceptual errors regarding the context of the interaction (one-shot vs. repeated; anonymous vs. public) might have selected for strategies that minimize the chance of making costly behavioral errors. Second, helping strangers might be a successful strategy for identifying other cooperative individuals in the population, where partner choice can turn strangers into interaction partners. Third, in contrast to the assumptions of the prisoner's dilemma model, it is possible that benefits of cooperation follow a non-linear function of investment. Non-linear benefits result in negative frequency dependence even in one-shot games. Finally, in many real-world situations individuals are able to parcel investments such that a one-shot interaction is turned into a repeated game of many decisions.

Keywords: human cooperation, one-shot games, prisoner's dilemma, error-management, cultural group selection, strong reciprocity

HUMAN HELPFULNESS

One doesn't have to observe humans in their natural habitat for long to witness many and varied examples of prosocial behavior, often directed towards complete strangers. People might vacate a seat on a crowded bus or train to let an elderly person sit down; hold open a door for others; or help a struggling parent to carry their pram down a flight of stairs. Humans also willingly donate resources, such as money or food, to others for example by giving to charity (e.g., Frey and Meier, 2004; Soetevent, 2005). These charitable donations are typically directed at recipients the helpful individual does not know or will expect to meet in the future. This propensity to help unrelated others who reside outside our regular social circle is striking when one considers that these helpful acts are seemingly unobserved and many of the interactions are unlikely to persist beyond the current round. To explain why humans are so helpful under conditions that seem to predict selfishness, researchers have attempted to bring this behavior into the laboratory by creating stylized games where

subjects can cooperate with or help one another, but where payoffs are structured such that self-interested behaviors yield greater rewards. One of the most widely-used paradigms is the anonymous one-shot prisoner's dilemma game (Luce and Raiffa, 1957). In this game, two interacting players simultaneously choose between cooperating or defecting. Payoffs are structured such that mutual cooperation yields higher payoffs for both than mutual defection, but defecting yields a higher individual payoff than cooperating regardless of how the partner behaves. Hence there is a temptation to defect even if the partner cooperates (see **Figure 1**). This game can be modeled with more than two players with similar incentives, yielding a temptation to defect when others cooperate: this is an n-player prisoner's dilemma game (also commonly referred to as a public goods game, Ledyard, 1995). Regardless of whether the game is played with two or more than two players, in a one-shot prisoner's dilemma the evolutionarily stable strategy (Maynard-Smith, 1982) is to defect. Another game that has been widely used is the dictator game (Kahneman et al., 1986).

Player 1→ Player 2 ↓	C	D
C	R, R	T, S
D	S, T	P, P

FIGURE 1 | Payoffs accruing to (Player 1, Player 2) according to each player's decision to cooperate (C) or defect (D) in a social dilemma are shown. R is the reward for mutual cooperation, T is the temptation to defect, S is the sucker's payoff and P is the punishment for mutual defection. A game satisfies the assumptions of the prisoner's dilemma where $T > R > P > S$. The snowdrift game is captured wherever $T > R > S > P$. Thus, the prisoner's dilemma and the snowdrift game differ only in the best possible response to a partner's defection: in the prisoner's dilemma, the best response is to defect whereas in the snowdrift game, the best response is to cooperate.

This is a two-player game where one player (the "dictator") is endowed with a sum of money and can choose to give none, some or all of the endowment to the partner. The "receiver" in the game has no power and must accept any division of the endowment offered by the dictator. As in the prisoner's dilemma game, individuals can maximize their earnings in the game by behaving selfishly and keeping the entire endowment.

Empirical studies have shown that humans often cooperate in anonymous one-shot prisoner's dilemma games (or give money in dictator games) despite the fact that not contributing is the optimal solution (Camerer, 2003; Henrich et al., 2005; Engel, 2011). It has been suggested that helping in the absence of obvious rewards emerges from proximate psychological mechanisms that make helping others subjectively rewarding for the actor (Andreoni, 1990; Fehr and Camerer, 2007; Harbaugh et al., 2007). To explain why these psychological mechanisms exist, one has to ask under which conditions they are likely to have been favored by Natural Selection. The two dominant current explanations—the "big mistake hypothesis" (also known as the "mismatch hypothesis") and the "cultural group selection hypothesis"—mainly disagree on the question of whether anonymous one-shot interactions were absent in human evolutionary history until very recently, and hence whether helping strangers is based on a psychological constraint or is instead adaptive under certain circumstances (specifically, when interaction partners belong to the same cultural group) (Rand and Nowak, 2013). Here, we first present these two hypotheses in more detail. We then move on to discuss how the laboratory game settings may differ from real-life interactions with strangers not only with respect to population structure but also in other fundamental ways. Based on this assessment, we will argue that there are several plausible routes by which cooperative behavior towards strangers could increase an individual's direct fitness (**Table 1**). We hope that this evaluation will stimulate researchers to use or develop new experimental paradigms, such that our knowledge of the proximate mechanisms that underpin cooperation can be complemented with a better understanding of the adaptive significance of such strategies.

COOPERATION—A BIG MISTAKE?

Some researchers have argued that the expression of prosocial behaviors in laboratory anonymous, one-shot encounters can be explained by mis-firing of psychological mechanisms in a context we are not evolutionarily adapted to (e.g., Cosmides and Tooby, 1989; Hoffman et al., 1998; Ben-Ner and Putterman, 2000; Johnson et al., 2003; Tooby et al., 2006). The line of reasoning is that the proximate mechanisms underpinning human social behavior evolved in close-knit, small groups comprised of kin and stable interaction partners, where no interaction was ever truly one-shot or anonymous. In such an environment, an individual that was predisposed to help group members would likely have been compensated for their investment, either via indirect benefits to kin (Hamilton, 1964a,b) or via personal benefits arising from future interactions with the recipient (Trivers, 1971) or observers (Nowak and Sigmund, 1998; Roberts, 1998). It is clear that these putative ancestral environments were fundamentally different to the modern day environment of most humans and especially to the artificial setup of most laboratory games. It is argued, therefore, that subjects in laboratory settings rely on psychological mechanisms that evolved in the context of repeated, non-anonymous interactions and that our predisposition to cooperate in the lab (or with strangers in the real world) might, therefore, be an evolutionary relic of times gone by. In support of the so-called (Boyd and Richerson, 2002; Fehr and Henrich, 2003) mismatch or big mistake hypothesis, there are several instances where imperfect human behavior can be attributed to the mismatch between our evolved psychology and our current environment. For example, the common tendency to be phobic of ancestrally dangerous stimuli (e.g., snakes, spiders) relative to more pressing modern-day threats (e.g., cars, guns); and the proclivity to ingest excessive quantities of foods that are high in sugar, fat and salt based on their relative abundance nowadays as compared with ancestral environments (Irons, 1998) can both be explained as psychological mechanisms that on average produced fitness enhancing outcomes in ancestral environments but that no longer deliver such benefits—and can even be harmful—in the modern world (Hagen and Hammerstein, 2006; Tooby et al., 2006).

Nevertheless, critics of the big mistake hypothesis have questioned the validity of assumption that ancestral social environments were devoid of anonymous one-shot encounters. Instead, it seems probable that even ancestral hunter-gatherers probably had some encounters with strangers with no obvious future follow-up, for example in the context of interactions outside their immediate social group (e.g., trading, feuds and so on, Lee, 1972; Wiessner, 1982; Keeley, 1996; Fehr and Henrich, 2003; Hill et al., 2011). It can be shown that, despite our

Table 1 | The key features of different theories to explain why humans cooperate in ostensibly anonymous, one-shot encounters.

Theory	Where did strategies used in lab games evolve?	How might individuals benefit from helping strangers?	Is helping behavior adaptive in anonymous, one-shot lab games?
Mismatch/ Big-mistake	Strategies evolved in context of repeated interactions with known partners.	No benefit	No
Between group competition	Strategies evolved to deal with (in-group) strangers even in one-shot, anonymous situations.	Benefit occurs in context of between-group competition. Unclear whether individual benefits directly (interdependence) or whether benefits accrue indirectly to kin residing in the group.	Maybe—if other people in the game are the relevant in-group and if helping in this setting somehow increases the group's performance in competition with other groups. This prediction needs empirical verification.
Misperceiving interaction duration or anonymity	Strategies evolved to deal with strangers in situations with a probability of being anonymous/one-shot but where there was uncertainty over these parameters.	Individuals benefit from helping strangers if there are asymmetries in costs of misperceiving interaction duration/anonymity.	No
Partner choice via exploration	Strategies evolved to identify potential cooperative interaction partners from the population.	Individuals benefit from helping if this allows them to identify cooperative partners in the population and to establish relationships with these partners.	No
Alternative payoff matrices	Strategies evolved in the context of non-linear games.	Individual can benefit from helping if benefits of investing in public good are negative frequency dependent.	No
Parceled investments	Strategies evolved in contexts where individuals could parcel investments into smaller units, thereby turning one-shot interactions into games involving repeated decisions.	Individuals benefit if partner's cooperative decisions/continuing the interaction are conditional upon the individual acting helpfully.	No

evolutionary heritage, humans are able to identify interactions where helping is likely to yield a return benefit (e.g., by identifying kin, by predicting when behavior is likely to be observed or not, or when interactions are likely to be repeated)—and adjust behavior accordingly (Fehr and Henrich, 2003). For example, various studies have shown that people recognize and preferentially help kin over non-kin (Barber, 1994; Gurven, 2004); that we cooperate more when investments are public rather than anonymous (e.g., Milinski et al., 2002; Andreoni and Petrie, 2004; Soetevent, 2005; Lamba and Mace, 2010); and that we cooperate less when interactions will not be repeated or will end soon (e.g., Gächter and Falk, 2002; Gächter et al., 2008). Moreover, recent work has suggested that while people may use heuristics from daily life to guide cooperative decision making, these heuristics can be rapidly updated to reflect the conditions imposed in the artificial lab setting (Rand et al., 2012, 2014). In fact even fish are apparently able to adjust levels of cooperation to the likelihood of repeated interactions (Oates et al., 2010). Thus, critics of the big mistake hypothesis have good arguments that humans do have the cognitive machinery to detect when an interaction is likely to yield direct return benefits, and to adjust behavior accordingly.

If, as seems likely, interactions with strangers probably did exist in our evolutionary history, why do humans have a psychology that seems geared towards cooperating in these contexts, given that the short-term, payoff-maximizing approach is apparently to defect under such conditions? For cooperative investments to come under positive selection, the behavior must form part of a strategy that on average increases the fitness of the bearer. In other words, cooperative actors must somehow ultimately be repaid for their investments. Broadly, it has been suggested that the ultimate benefits associated with making costly investments could arise either with or without assortative population structure (often referred to as group-level (or multi-level) and individual-level selection in the literature). We consider these two arguments in turn.

BETWEEN-GROUP COMPETITION
Whenever populations are structured into groups of relatively stable composition, there is potential for competition between groups. The stronger such competition between groups is the more individual and group interests are closely aligned, and selection may favor strategies that prioritize unconditional contributions towards group success. The tendency for such

"multi-level selection" to promote cooperation is seen at all levels of life and several of the major evolutionary transitions identified by Szathmáry and Smith (1995) rely on the repression of lower order conflict to bring about a higher-level advantage. According to accounts of helping based on between-group competition in humans (e.g., Gintis, 2000; Henrich, 2004; Boyd and Richerson, 2009), group-level benefits favor individual costly investments (even in the absence of obvious mechanisms to be directly compensated) because within-group helping generally predicts group success in the face of extinction threats (e.g., due to competition with other groups, environmental catastrophes etc., Gintis, 2000). Similar arguments emphasizing the importance of group-level benefits have been proposed by biologists for the evolution of helping behavior among non-relatives in cooperatively breeding groups (e.g., Kokko et al., 2001) and have been formulated with the notion of interdependence replacing the relatedness term in Hamilton's rule (Roberts, 2005). Although group-level benefits could theoretically arise via genetic group selection, the high levels of between-group genetic variance that would be necessary to facilitate selection are thought to be unrealistic given the genetic evidence for (female) migration among groups (Seielstad et al., 1998; Bell et al., 2009). Moreover, since genetic between-group selection is thought to be a small force in large groups, it is unlikely to account for large-scale cooperation seen in modern day human societies. Thus, colleagues have focussed instead on the concept of cultural group selection, whereby immigrating individuals are expected to adopt the cultures of the new group, thereby reducing the behavioral variance that migrating individuals would otherwise have (Boyd and Richerson, 1982; Henrich, 2004; Bell et al., 2009). Cultural group selection of cooperative traits is thought to be facilitated by a general predisposition to learn socially from others, which would be adaptive where the cost of information acquisition was sufficiently high (Boyd and Richerson, 1982, 2009; Henrich and Boyd, 2001; Guzmán et al., 2007; Richerson and Boyd, 2008; but see Eriksson and Coultas, 2009; André and Morin, 2011). Selection for social learning outside the cooperative domain could then facilitate the evolution of cooperative cultural norms within populations. If populations varied in these norms and if the outcomes of competition between populations varied according to within-group levels of cooperation, then cultural group selection could favor cooperative strategies. The benefits of within-group cooperation could, in turn, select for psychological predispositions to cooperate with in-group members while punishing defectors (i.e., "strong reciprocity", Gintis, 2000; Fehr et al., 2002; Fehr and Henrich, 2003; Boyd and Richerson, 2009), even in anonymous, one-shot encounters. Evidence for cross-cultural variation in cooperative tendency across societies (Henrich et al., 2001, 2005, 2006, 2010; Gächter and Herrmann, 2009; Gächter et al., 2010; House et al., 2013) has been touted as key evidence for the existence of between-group variation in cooperative norms that could be the basis for such selection (Henrich, 2004; Henrich et al., 2005).

Nevertheless, the assumptions underpinning the cultural group selection account of human cooperation have been subject to debate (Burnham and Johnson, 2005; Hagen and Hammerstein, 2006; West et al., 2011). For instance, while it is often assumed that the predisposition for conformity necessary to catalyze the emergence of within-group cooperation is an adaptive trait, it has been demonstrated that non-conformist transmission dominates conformist strategies in evolutionary models (Eriksson and Coultas, 2009). Moreover, analytical models have shown that a tendency for conformist transmission can undermine the evolution of helping behaviors (Feldman et al., 1985; Lehmann et al., 2008), not catalyze them as was originally predicted. Thus, in contrast to the predictions of the cultural group selection models, it is apparently easier for costly helping strategies to evolve under genetic systems of inheritance rather than via culture. Another central assumption of cultural group selection models is that individuals are constrained by a predisposition for conformity to blindly adopt the behaviors of others, even when strategic non-conformity would increase biological fitness (e.g., Gintis, 2003; Boyd and Richerson, 2009). The validity of this assumption is contested (Hagen and Hammerstein, 2006; André and Morin, 2011; El Mouden et al., 2014; Morin, 2014). Instead, it has been argued that the conditions under which cultural group selection would be expected to produce a maladaptive tendency to copy altruistic behavior are prohibitively restrictive (El Mouden et al., 2014; Morin, 2014) and, moreover, that antagonistic co-evolution should act to curb psychological predispositions to copy maladaptive actions (El Mouden et al., 2014). Indeed, strong empirical evidence suggests that people are sensitive to the benefits of imitation and do not copy blindly as supposed (Rendell et al., 2011; Morgan et al., 2012; Morin, 2014). More recently, it has also been suggested that asymmetries among individuals within groups can facilitate within-group cooperation when there is between-group competition, without having to resort to cultural transmission, reciprocity or punishment (Gavrilets and Fortunato, 2014). Specifically, when some individuals are able to benefit more than others from the production of a (within-group) public good (e.g., if they are dominant to others in the group), it pays for these individuals to invest more in the production of the public good, and this effect is particularly pronounced when groups are in competition with one another. Thus, population structure and between-group competition can facilitate within-group cooperation even in the absence of culturally transmitted tendencies to copy cooperative and punitive behaviors.

Empirical evidence for cultural group selection has also been called into question. While the existence of inter-cultural variation in cooperative norms (e.g., Henrich et al., 2001, 2005) seemed initially supportive of the predictions of the cultural group selection models, more recent empirical work has shown that substantial within-culture variation in cooperation exists that can be explained by local demography and ecology rather than culture (Lamba and Mace, 2011; Nettle et al., 2011; Schroeder et al., 2014). The extent of within-culture variation has been demonstrated to be comparable to that previously observed between cultures (Lamba and Mace, 2011) and, since many of the former studies sampled only one or a few populations per culture, it is argued that much of the measured variation that has been attributed to cultural differences may not in

fact exceed within-group variation in these traits. Furthermore, cultural group selection models do not predict unconditional help towards any recipient but instead only towards members of the relevant in-group. Out-group members should not receive help and may in fact be harmed ("parochial altruism", Bernhard et al., 2006). There is no a-priori reason for human subjects in laboratory one-shot games to assume that the co-players are members of the in-group, and so deserving of help. One might just as easily expect that subjects assume that co-players are out-group members, which would not predict costly cooperative behavior. Finally, it appears biologically questionable that in-group members are indeed unfamiliar with each other and that they can be reasonably certain that there will not be any future interactions; such a scenario would be much more likely with out-group members. Thus, even with a cultural group selection account of cooperation, human behavior in stylized laboratory games still remains a puzzle because we have to understand why co-players are apparently treated as in-group rather than out-group members (Burnham and Johnson, 2005).

DIRECT BENEFITS WITHOUT BETWEEN-GROUP COMPETITION

The cultural group selection approach makes assumptions about population structure (grouping) and competition between these units without specifying whether a tendency to help strangers increases indirect fitness (through relatedness) or direct fitness (through interdependence). An alternative approach is to investigate conditions under which helping strangers may yield direct benefits in the absence of any specific population structure. One plausible suggestion is that human cooperation in one-shot games can be thought of as a strategy that minimizes costly error types (Burnham and Johnson, 2005; Yamagishi et al., 2007; Delton et al., 2011; Morin, 2014). Error management theory assumes that where there is uncertainty over the perceptual accuracy of the environment (such that all is not necessarily as it seems) and there are asymmetries in the costs of false-positive and false-negative error types, then evolution should favor strategies that minimize the costlier of the two error types (Johnson et al., 2013). It is important to bear in mind that Natural Selection is expected to favor strategies that on average increase the fitness of the bearer rather than to produce perfect behavior in every context. In other words, assessment errors mean that adaptive strategies will sometimes produce behavioral mistakes (West et al., 2011; Morin, 2014). For example, consider a meerkat who hears an alarm call indicating the likely, but only probabilistic, presence of a predator. In such a scenario there are likely to be asymmetric costs associated with erroneous behavioral responses. Responding to a false alarm call by fleeing incurs energetic and opportunity costs, while failing to flee when the threat is real incurs a far higher possible cost of being caught by a predator. Based on these asymmetric costs of behavioral errors, a strategy of "if hear alarm call, then flee" might be on average adaptive even if it produces several behavioral errors (i.e., fleeing in response to false alarms). In the context of interactions with strangers, it may well be the case that humans experience perceptual uncertainty over several features of the

interaction, any or all of which may favor strategies that err on the side of caution by cooperating even when there is little ostensible benefit to doing so. The uncertainty could stem from inaccuracies in perceiving the likely duration of the interaction, whether the interaction is truly anonymous, or the payoff matrix of the interaction. We discuss how perceptual uncertainty in any of these features might select for broadly cooperative strategies below.

MISPERCEIVING INTERACTION DURATION OR ANONYMITY

It has been argued that cooperation can be favored by evolution if there is even a small possibility that the interaction will be repeated and if this possibility is fundamentally unpredictable (Delton et al., 2011). Such a strategy could be adaptive even if it produces several behavioral errors (i.e., cooperating when no return benefits are possible). In a laboratory setting, an experimenter can exogenously impose the one-shot structure on the game (such that subjects can be certain that the interaction is not repeated) but this is unlike real life interactions with strangers, where we might often experience a degree of uncertainty about whether we might meet again. Using agent-based simulations, Delton et al. (2011) showed, using a mix of agents playing either always-defect (ALLD) or the conditionally cooperative strategy tit-for-tat (TFT), that cooperation can indeed be favored so long as there is a non-zero probability that the interaction might be repeated (and the partner is TFT rather than ALLD). In this setting, uncertainty over the number of interactions favors cooperation also in interactions that turn out to be one-shot. Cooperation is favored because mistaking a repeated interaction for one-shot (and therefore defecting in the first round of the game) incurred a greater cost than mistaking a one-shot interaction for repeated (and therefore erroneously cooperating). This is due to the unforgiving nature of TFT, whereby defecting in the first round of the game prompts the partner to defect in the next round and thus establishes mutual defection for the duration of the interaction. Strategies that take a chance on the interaction being repeated (and the partner playing TFT) by cooperating in the first round could instead establish mutually productive, cooperative relationships with TFT partners. In support of the idea, it has additionally been argued that due to autocorrelation of individual locations over time, interacting with an individual once leads to an increased probability of interacting with the same individual again in the future (Krasnow et al., 2013). Thus, by definition, meeting a partner once implies that the interaction will be repeated and that conditionally cooperative strategies will prevail.

Nevertheless, the generality of these findings have been challenged on the grounds that only interactions with unrepentant (ALLD) and unforgiving (TFT) strategies were explored (McNally and Tanner, 2011; Zefferman, 2014a,b). ALLD is unrepentant in the sense that it is committed to play defect in all rounds; it cannot change its behavior if the interaction turns out to be repeated. TFT is unforgiving in the sense that any defection by the partner will be immediately reciprocated with defection. By contrast, it was verbally argued that the importance of uncertainty for catalyzing one-shot cooperation

might be substantially reduced if agents employed strategies that allowed for flexible responses, either an increased propensity to cooperate once the interaction extended beyond round one or a non-zero probability to forgive cheating partners in a repeated interaction (McNally and Tanner, 2011). This is because, with flexible strategies, the importance of cooperating in the first round would be reduced substantially since cooperation could still be established (or re-established) beyond round one (McNally and Tanner, 2011). Thus, the question of why individuals cooperate in ostensible one-shot interactions would remain unresolved. Indeed, the inclusion of repentant and forgiving strategies under the same conditions of uncertainty has subsequently been shown to vastly diminish the advantage to cooperating in (ostensible) one-shot interactions; in some cases actually reversing the direction of selection (Zefferman, 2014a). Zefferman (2014a) proposes that the failure of the model to predict one-shot cooperation when an arguably more realistic strategy set is considered emphasizes the importance of social learning of local cooperative norms (i.e., the cultural group selection approach) for explaining one-shot cooperation. Ultimately, empirical studies are likely to be important for understanding whether defection in an ostensible one-shot encounter precludes cooperation from being established (as predicted by Delton et al., 2011; Delton and Krasnow, 2014) or, instead, whether humans are more likely to forgive an interaction partner who starts by defecting but then switches to cooperation if the interaction continues.

While the Delton et al. (2011) model assumed that psychological responses to cooperate evolved in the context of directly reciprocal interactions, one also has to consider that under real world scenarios, actions might also be observed by uninvolved bystanders, who adjust their behavior towards the actor accordingly. For example, under indirect reciprocity models, helpful acts are reciprocated by third-parties rather than by recipients. Misperceiving that an interaction is unobserved by bystanders might carry similar costs to misperceiving the likely duration, in that erroneous defection incurs greater costs than erroneous cooperation in both scenarios. Thus, error-management might still play a role in sustaining cooperation but because individuals can never be certain that their actions are unobserved, rather than because they misperceive interaction duration. If there is even a slight possibility that actions will be observed—and if being seen as unhelpful carries greater costs than helping when no one is watching (as has been proposed in models of judgment bias, Rankin and Eggimann, 2009)—then selection might favor psychological mechanisms that make us behave as though we are observed most of the time. Empirical evidence suggests that reputation concerns have an important influence on prosocial tendency: people are typically more cooperative in public rather than anonymous laboratory games (e.g., Andreoni and Petrie, 2004; Lamba and Mace, 2010) and even exposing people to subtle cues of being watched (in the form of eye images) increases prosocial behavior under some circumstances (e.g., Haley and Fessler, 2005; Bateson et al., 2006; but see Fehr and Schneider, 2010; Raihani and Bshary, 2012 for failed replications). The presence of potential observers is made even more important

when one considers that, via gossip, one's positive or negative actions could be broadcast to several "observers", who need not even have been present at the time of the event (e.g., Sommerfeld et al., 2007, 2008; Feinberg et al., 2014). Thus, an error-managing strategy might often cooperate—even when interactions seem to be anonymous—to minimize the reputation costs of not cooperating if the interaction turns out to be observed.

PARTNER CHOICE VIA EXPLORATION

Many of the games used to explore cooperation under laboratory settings impose a forced-play structure on subjects: players cannot choose who they want to interact with or to leave unproductive relationships (Axelrod and Hamilton, 1981). While cooperation can evolve under such circumstances in artificial simulations (e.g., via clustering or assortment of cooperators, Nowak and May, 1992; Fletcher and Doebeli, 2009; but see Hauert and Doebeli, 2004), network reciprocity based on spatial structure does not seem to support cooperation in empirical studies (Grujić et al., 2014). Indeed, assuming that players are constrained to use pure strategies and are forced to interact with one another is unlikely to reflect the conditions under which real-word relationships operate. Instead, individuals are typically able to choose interaction partners, and can choose to continue interactions with cooperative partners while terminating relationships that prove unproductive (e.g., Noë and Hammerstein, 1994; Baumard et al., 2013).

The possibility for partner choice might therefore favor unconditionally helpful strategies if being observed as helpful increases the chance of being chosen for future interactions. Importantly, the chooser need not make costly investments to reimburse the helper for their actions. Instead, simply being chosen for a mutually productive interaction (e.g., producing offspring, cohabiting) could compensate the helper for their initial investment (e.g., Bshary and Grutter, 2006; McNamara et al., 2008). According to the "competitive altruism" theory, competition for interaction partners occurs within a biological market (Noë and Hammerstein, 1994); and individuals who produce the strongest signals of quality (via helping) will be preferred as partners (Roberts, 1998; Lotem et al., 2003; Barclay, 2011). Empirical evidence suggests that competitive altruism might be an important mechanism underpinning human helping behavior: people choose interaction partners based on cooperative reputation (e.g., Barclay and Willer, 2007; Sylwester and Roberts, 2010, 2013) and avoid defectors (Rockenbach and Milinski, 2011), hence individuals compete with one another to advertise helpful actions (Raihani and Smith, in press).

The possibility for partner choice could also promote cooperation, even with unknown individuals in an anonymous setting, because individuals can use cooperative first moves to test the response of the partner and then decide whether to continue the interaction or not. Error-management might still play a role but, unlike the Delton et al. (2011) model, the cost of not cooperating would be that one misses out on the chance to have a mutually productive relationship with the partner, rather than that one is stuck in a mutually destructive

relationship. It has been demonstrated that where there is extrinsically maintained variation in cooperative tendency within a population (maintained by differences in quality or ability to invest in the partner, McNamara and Leimar, 2010, or by mutation, immigration, recombination or epistasis, McNamara et al., 2004), then this variation could select for cooperative strategies because this is a way to identify whether the partner is also cooperative (McNamara et al., 2004). In such scenarios cooperation can evolve whenever the benefit of interacting with a cooperator outweighs the benefit of occasional exploitation. Variability in cooperativeness, together with a long lifespan during which to reap the benefits of a productive partnership, can then pave the way for the evolution of choosiness because, given sufficient variation in partner quality it can pay to leave a less cooperative partner in hope of finding a more cooperative individual next time (Sherratt and Roberts, 1998; McNamara et al., 2008). In many real-world scenarios, initial cooperative acts might often be low cost (in comparison to the potential benefits of establishing a mutually productive relationship) but investments could increase as the relationship becomes more established (e.g., Roberts and Sherratt, 1998). For example, while we do not routinely see people handing out $100 bills to strangers, low cost helpful acts, such as holding a door open or assisting with heavy bags, are relatively commonplace. These low cost investments are consistent with the idea that cooperation could be used as an exploratory strategy to strike up mutually productive relationships with other individuals in the population. Nowadays, modern technology may even allow us to develop long-distance relationships as a consequence of chance encounters where we help or are helped by others.

ALTERNATIVE PAYOFF MATRICES

Most theoretical and laboratory studies of cooperation in humans have assumed a prisoner's dilemma type payoff-matrix, where benefits scale linearly with investments. In such scenarios, the payoff-maximizing strategy in a one-shot game is to defect regardless of how the partner(s) behave. As a consequence, explanations based on assortment, repeated interactions or relatedness are typically invoked to account for the emergence and stability of cooperative behavior. The assumption that all social dilemmas have the structure of an n-player prisoner's dilemma is, however, flawed (Kollock, 1998; Archetti and Scheuring, 2011). Alternative social dilemmas with different payoff matrices can yield evolutionarily stable cooperative strategies without having to invoke assortment, relatedness or repeated interactions (e.g., Doebeli and Hauert, 2005; Archetti, 2009a,b; Archetti and Scheuring, 2011). For example, consider the 2-player snowdrift game. This game describes two drivers traveling in opposite directions when they come across a snowdrift blocking the road. Neither driver can get home unless the road is cleared. Although each driver prefers the other to do the clearing, each would also rather clear the snowdrift himself than for the snowdrift to not be cleared at all. As a consequence, in the two-player snowdrift game, the best response to a cooperative partner is to defect, while the best response to a defecting partner is to cooperate. Thus, the strategic payoffs differ markedly from those in more frequently used prisoner's

dilemma, where defecting always yields the highest payoff (Doebeli and Hauert, 2005; **Figure 1**). An n-player snowdrift game is often referred to as a volunteer's dilemma (Diekmann, 1985). In its simplest form, this game assumes that a public good will be produced if one player cooperates and that additional investments do not increase the magnitude of the public good. Thus, unlike the traditional n-player prisoner's dilemma, benefits are a non-linear function of investment and cooperation is therefore expected to be under negatively frequency dependent selection. As in the snowdrift game, the benefit of the public good being produced is larger than the cost associated with producing it, such that all players would do best to invest to produce the public good if no one else does so.

Relaxing the assumption of linearity has far-reaching consequences for the emergence and stability of cooperative strategies in n-player games. Specifically, when individuals are unsure about how others are likely to behave in a non-linear public goods game, then the best strategy is to cooperate probabilistically (where the probability depends on the cost to benefit ratio of cooperating and group size, Archetti, 2009a,b; Archetti and Scheuring, 2011, 2012). Such probabilistic strategies will emerge even in non-repeated games without spatial assortment or interactions among relatives. Where players differ in their ability to invest, or in the benefit that they can extract from the public good being produced, then this can also offer a potential solution to a non-linear public goods game, with those players who will reap the largest benefit from investing being more likely to contribute to the public good (e.g., see Gavrilets and Fortunato, 2014; Szolnoki and Perc, 2014). With respect to the assumption of linearity in benefits, n-player prisoner's dilemma games lie at one end of a spectrum with threshold Public Goods Games (i.e., volunteer's dilemmas) at the other end. All intermediate cases (where benefits are a sigmoidal function of investment in the public good) resemble the volunteer's dilemma more than the traditional n-player prisoner's dilemma in that they also yield a stable mixed equilibrium of cooperators and defectors in the population, even in the absence of other incentives to cooperate (Archetti and Scheuring, 2011). Thus, it has been argued that many biological examples of cooperation in social dilemmas are more likely to yield non-linear rather than linear benefits, which has profound implications for our understanding of how cooperation evolves and is maintained in these scenarios (Kummerli et al., 2007; Archetti, 2009a; Sherratt et al., 2009; Archetti et al., 2011; Archetti and Scheuring, 2012). For example, the costly production of invertase in yeast, alarm calling in animal groups and the formation of fruiting bodies in social amoebas are all examples that can be described as non-linear public goods games where cooperation is under negative frequency dependent selection (Gore et al., 2009; Archetti and Scheuring, 2011; Archetti et al., 2011). Yeast growth requires the costly production of the enzyme invertase to hydrolyze sucrose into smaller glucose molecules which can be imported into the cell (Gore et al., 2009). Although invertase production is costly and can be parasitized by non-producing cells, a complete lack of invertase can be lethal, meaning that producer cells outperform non-producers when rare. Conversely, at high densities of

producers, non-producing cells have an advantage because they can parasitize the invertase being produced by the other cells (Gore et al., 2009).

For humans, it is less clear whether the majority of the social dilemmas that have shaped our social behavior ought to be described with linear or with non-linear payoff functions. In the case of punishment, which has been modeled as a second-order public good (and often therefore explained in terms of cultural group selection, e.g., Boyd et al., 2003), it has been argued that the payoffs of investing in punishment (in terms of increased within-group cooperation) are likely to scale non-linearly with number of punishers, thereby providing a direct individual-level solution for its existence (Raihani and Bshary, 2011). Other social dilemmas that have been explained in terms of cultural group selection, for example contributions to group defense during war, might also be more likely to have non-linear than linear payoffs. Group survival, which is the typical currency for payoffs associated with cooperating in warfare, is likely to be a non-linear function of contributions to defense, meaning that the payoffs associated with increasing within-group cooperation are by definition non-linear. It may turn out to be the case that linear public goods problems exist mainly in artificial laboratory settings and that subjects use strategies and psychology that evolved predominantly in the context of non-linear games when they participate. Specifically, if most real-world public goods problems are non-linear in nature and if there is always a certain degree of uncertainty about whether others will contribute to produce the public good (thereby obviating the need for the subject's own investment), selection may have favored strategies that either probabilistically cooperate (when cooperation is binary) or that invest intermediate amounts (when cooperation is a continuous variable) (c.f. Kummerli et al., 2010), even in one-shot games.

PARCELED INVESTMENTS

Finally, we would like to highlight an additional discrepancy with the way laboratory prisoner's dilemma experiments are set up compared with how interactions typically occur in the real-world. In many experimental games, the act to cooperate or to defect is an all-or-nothing action where players press a button, and typically learn about each other's choices *post hoc*. On the other hand, interactions with strangers in our evolutionary past (e.g., in the context of trades) are highly unlikely to have involved exchange of closed boxes where players only found out after separating what the other put in the box. Instead, real-world interactions with strangers might often have involved simultaneous or parceled exchanges, where individuals could monitor the behavior of one another, make behavioral adjustments in real time and—importantly—terminate unproductive exchanges prematurely (e.g., Connor, 1992; Hart and Hart, 1992). For example, most female lions approach intruders simulated by playbacks more slowly if they teamed up with laggard female group members, apparently looking back regularly to check the spatial configuration of self vs. partners (Heinsohn and Packer, 1995). Theoretical models have shown that where cooperation is not an all-or-nothing event but investments can instead can be parceled and

adjusted in real time, then prisoner's dilemma type situations can be solved cooperatively, even in one-shot anonymous games (Friedman and Hammerstein, 1991; Connor, 1995; Johnstone and Bshary, 2002). The key issue is that stable cooperation does not rely on repeated interactions but on repeated decisions. The question of whether encounters between strangers (either currently, or in our evolutionary past) are more likely to involve single vs. multiple decisions remains open for empirical exploration.

CONCLUSION

We have discussed potential explanations for the observation that humans help complete strangers under natural conditions and often cooperate in laboratory anonymous one-shot games. While the extent to which humans encountered such situations in our evolutionary past necessarily remains an open debate, it is clear that such encounters happen today. Moreover, it appears that we are at least partly adapted to adjust behavior to such situations. Our aim was to identify functional explanations for why humans regularly help strangers under natural conditions. The main take-home message is that the laboratory settings deviate from natural encounters in various important ways that make helping in real-world encounters potentially under positive selection while it is clearly not in the lab setting. Error-management arguments suggest that during natural encounters it is very hard to assess whether an interaction will be one-off and/or whether the interaction will remain truly anonymous. Furthermore, while lab games typically impose a forced-play scenario on subjects, in real-life individuals can choose to pursue productive relationships and abandon unproductive partners. This possibility for partner choice might select for helping behavior even with unknown strangers. Another issue is that in real-world interactions, the payoffs might often be a non-linear function of total investment, causing helping to be under negative frequency dependent selection rather than being altruistic. Finally, as soon as interactions involve multiple decisions, stable cooperation may be achieved even between strangers without any future perspective. We note that these different explanations are not mutually exclusive and that different explanations are likely to apply to different real-world scenarios. One way to explore the importance of the different explanations we proposed would be to use a wider variety of laboratory games, that replicate conditions that are likely to prevail in real-world interactions with strangers.

REFERENCES

André, J.-B., and Morin, O. (2011). Questioning the cultural evolution of altruism. *J. Evol. Biol.* 24, 2531–2542. doi: 10.1111/j.1420-9101.2011.02398.x

Andreoni, J. R. (1990). Impure altruism and donations to public goods: a theory of warm-glow giving. *Econ. J.* 100, 464–477. doi: 10.2307/2234133

Andreoni, J., and Petrie, R. (2004). Public goods experiments without confidentiality: a glimpse into fund-raising. *J. Pub. Econ.* 88, 1605–1623. doi: 10.1016/s0047-2727(03)00040-9

Archetti, M. (2009a). Cooperation as a volunteer's dilemma and the strategy of conflict in public goods games. *J. Evol. Biol.* 22, 2192–2200. doi: 10.1111/j.1420-9101.2009.01835.x

Archetti, M. (2009b). The volunteer's dilemma and the optimal size of a social group. *J. Theor. Biol.* 261, 475–480. doi: 10.1016/j.jtbi.2009.08.018

Archetti, M., and Scheuring, I. (2011). Coexistence of cooperation and defection in public goods games. *Evolution* 65, 1140–1148. doi: 10.1111/j.1558-5646.2010.01185.x

Archetti, M., and Scheuring, I. (2012). Game theory of public goods in one-shot social dilemmas without assortment. *J. Theor. Biol.* 299, 9–20. doi: 10.1016/j.jtbi.2011.06.018

Archetti, M., Scheuring, I., Hoffman, M., Frederickson, M. E., Pierce, N. E., and Yu, D. W. (2011). Economic game theory for mutualism and cooperation. *Ecol. Lett.* 14, 1300–1312. doi: 10.1111/j.1461-0248.2011.01697.x

Axelrod, R., and Hamilton, W. (1981). The evolution of cooperation. *Science* 211, 1390–1396. doi: 10.1126/science.7466396

Barber, N. (1994). Machiavellianism and altruism: effect of relatedness of target person on Machiavellian and helping attitudes. *Psychol. Rep.* 75, 403–422. doi: 10.2466/pr0.1994.75.1.403

Barclay, P. (2011). Competitive helping increases with the size of biological markets and invades defection. *J. Theor. Biol.* 281, 47–55. doi: 10.1016/j.jtbi.2011.04.023

Barclay, P., and Willer, R. (2007). Partner choice creates competitive altruism in humans. *Proc. Biol. Sci.* 274, 749–753. doi: 10.1098/rspb.2006.0209

Bateson, M., Nettle, D., and Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* 2, 412–414. doi: 10.1098/rsbl.2006.0509

Baumard, N., André, J.-B., and Sperber, D. (2013). A mutualistic approach to morality: the evolution of fairness by partner choice. *Behav. Brain Sci.* 36, 59–78. doi: 10.1017/S0140525x11002202

Bell, A. V., Richerson, P. J., and McElreath, R. (2009). Culture rather than genes provides greater scope for the evolution of large-scale human prosociality. *Proc. Natl. Acad. Sci. U S A* 106, 17671–17674. doi: 10.1073/pnas.0903232106

Ben-Ner, A., and Putterman, L. (2000). On some implications of evolutionary psychology for the study of preferences and institutions. *J. Econ. Behav. Organ.* 43, 91–99. doi: 10.1016/s0167-2681(00)00110-4

Bernhard, H., Fischbacher, U., and Fehr, E. (2006). Parochial altruism in humans. *Nature* 442, 912–915. doi: 10.1038/nature04981

Boyd, R., Gintis, H. M., Bowles, S., and Richerson, P. J. (2003). The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. U S A* 100, 3531–3535. doi: 10.1073/pnas.0630443100

Boyd, R., and Richerson, P. J. (1982). Cultural transmission and the evolution of cooperative behavior. *Hum. Ecol.* 3, 325–351. doi: 10.1007/bf01531189

Boyd, R., and Richerson, P. J. (2002). Group beneficial norms can spread rapidly in a structured population. *J. Theor. Biol.* 215, 287–296. doi: 10.1006/jtbi.2001.2515

Boyd, R., and Richerson, P. (2009). Culture and the evolution of human cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 3281–3288. doi: 10.1098/rstb.2009.0134

Bshary, R., and Grutter, A. S. (2006). Image scoring and cooperation in a cleaner fish mutualism. *Nature* 441, 975–978. doi: 10.1038/nature04755

Burnham, T. C., and Johnson, D. D. P. (2005). The biological and evolutionary logic of human cooperation. *Anal. Kritik* 27, 113–135.

Camerer, C. (2003). *Behavioral Game Theory: Experiments in Strategic Interaction.* Princeton, New Jersey: Princeton University Press.

Connor, R. C. (1992). Egg-trading in simultaneous hermaphrodites: an alternative to Tit-for-Tat. *J. Evol. Biol.* 5, 523–528. doi: 10.1046/j.1420-9101.1992.5030523.x

Connor, R. C. (1995). Impala allogrooming and the parcelling model of reciprocity. *Anim. Behav.* 49, 528–530. doi: 10.1006/anbe.1995.0070

Cosmides, L., and Tooby, J. L. (1989). Evolutionary psychology and the generation of culture, part II: case study: a computational theory of social exchange. *Ethol. Sociobiol.* 10, 51–97. doi: 10.1016/0162-3095(89)90013-7

Delton, A. W., and Krasnow, M. M. (2014). An independent replication that the evolution of direct reciprocity under uncertainty explains one-shot cooperation: commentary on Zeffermann. *Evol. Hum. Behav.* 35, 547–548. doi: 10.1016/j.evolhumbehav.2014.07.004

Delton, A. W., Krasnow, M., Cosmides, L., and Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proc. Natl. Acad. Sci. U S A* 108, 13335–13340. doi: 10.1073/pnas.1102131108

Diekmann, A. (1985). Volunteer's dilemma. *J. Confl. Resol.* 29, 605–610. doi: 10.1177/0022002785029004003

Doebeli, M., and Hauert, C. (2005). Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecol. Lett.* 8, 748–766. doi: 10.1111/j.1461-0248.2005.00773.x

El Mouden, C., André, J.-B., Morin, O., and Nettle, D. (2014). Cultural transmission and the evolution of human behaviour: a general approach based on the Price equation. *J. Evol. Biol.* 27, 231–241. doi: 10.1111/jeb.12296

Engel, C. (2011). Dictator games: a meta-study. *Exp. Econ.* 14, 583–610. doi: 10.1007/s10683-011-9283-7

Eriksson, K., and Coultas, J. C. (2009). Are people really conformist-biased? an empirical test and a new mathematical model. *J. Evol. Psych.* 7, 5–21. doi: 10.1556/jep.7.2009.1.3

Fehr, E., and Camerer, C. F. (2007). Social neuroeconomics: the neural circuitry of social preferences. *Trends Cogn. Sci.* 11, 419–427. doi: 10.1016/j.tics.2007.09.002

Fehr, E., Fischbacher, U., and Gächter, S. (2002). Strong reciprocity, human cooperation and the enforcement of social norms. *Hum. Nat.* 13, 1–25. doi: 10.1007/s12110-002-1012-7

Fehr, E., and Henrich, J. (2003). "Is strong reciprocity a maladaptation?," in *Genetic and Cultural Evolution of Cooperation*, ed P. Hammerstein (Cambridge, MA: MIT Press), 55–82.

Fehr, E., and Schneider, F. (2010). Eyes are on us, but nobody care: are eye cues relevant for strong reciprocity? *Proc. Biol. Sci.* 277, 1315–1323. doi: 10.1098/rspb.2009.1900

Feinberg, M., Willer, R., and Schultz, M. (2014). Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* 25, 656–664. doi: 10.1177/0956797613510184

Feldman, M. W., Cavalli-Sforza, L., and Peck, J. L. (1985). Gene-culture coevolution: models for the evolution of altruism with cultural transmission. *Proc. Natl. Acad. Sci. U S A* 82, 5814–5818. doi: 10.1073/pnas.82.17.5814

Fletcher, J. A., and Doebeli, M. (2009). A simple and general explanation for the evolution of altruism. *Proc. Biol. Sci.* 276, 13–19. doi: 10.1098/rspb.2008.0829

Frey, B. S., and Meier, S. (2004). Pro-social behavior in a natural setting. *J. Econ. Behav. Organ.* 54, 65–88. doi: 10.1016/j.jebo.2003.10.001

Friedman, J. W., and Hammerstein, P. (1991). "To trade, or not to trade; that is the question," in *Game Equilibrium Models I: Evolution and Game Dynamics*, ed R. Selten (Berlin, Heidelberg: Springer), 257–275.

Gächter, S., and Falk, A. (2002). Reputation and reciprocity: consequences for the labor relation. *Scand. J. Econ.* 104, 1–26. doi: 10.1111/1467-9442.00269

Gächter, S., and Herrmann, B. (2009). Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. *Philos. Trans. R. Soc. Lond. B Sci.* 364, 791–806. doi: 10.1098/rstb.2008.0275

Gächter, S., Herrmann, B., and Thöni, C. (2010). Culture and cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2651–2661. doi: 10.1098/rstb.2010.0135

Gächter, S., Renner, E., and Sefton, M. (2008). The long-run benefits of punishment. *Science* 322:1510. doi: 10.1126/science.1164744

Gavrilets, S., and Fortunato, L. (2014). A solution to the collective action problem in between-group conflict with within-group inequality. *Nat. Commun.* 5:3526. doi: 10.1038/ncomms4526

Gintis, H. (2000). Strong reciprocity and human sociality. *J. Theor. Biol.* 206, 169–179. doi: 10.1006/jtbi.2000.2111

Gintis, H. (2003). The hitchhiker's guide to altruism: gene-culture coevolution and the internalization of norms. *J. Theor. Biol.* 220, 407–418. doi: 10.1006/jtbi.2003.3104

Gore, J., Youk, H., and can Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature* 459, 253–256. doi: 10.1038/nature07921

Grujić, J., Gracia-Lázaro, C., Milinski, M., Semmann, D., Traulsen, A., Cuesta, J. A., et al. (2014). A comparative analysis of spatial Prisoner's Dilemma experiments: conditional cooperation and payoff irrelevance. *Sci. Rep.* 4:4615. doi: 10.1038/srep04615

Gurven, M. (2004). To give and to give not: the behavioral ecology of human food transfers. *Behav. Brain Sci.* 27, 543–559. doi: 10.1017/s0140525x04000123

Guzmán, R. A., Rodríguez-Sickert, C., and Rowthorn, R. (2007). When in Rome, do as the Romans do: the coevolution of altruistic punishment, conformist learning and cooperation. *Evol. Hum. Behav.* 28, 112–117. doi: 10.1016/j.evolhumbehav.2006.08.002

Hagen, E. H., and Hammerstein, P. (2006). Game theory and human evolution: a critique of some recent interpretations of experimental games. *Theor. Popul. Biol.* 69, 339–348. doi: 10.1016/j.tpb.2005.09.005

Haley, K. J., and Fessler, D. M. T. (2005). Nobody's watching? Subtle cues affect generosity in a one-shot game. *Evol. Hum. Behav.* 26, 245–256. doi: 10.1016/j.evolhumbehav.2005.01.002

Hamilton, W. (1964a). The genetical evolution of social behaviour I. *J. Theor. Biol.* 7, 1–16. doi: 10.1016/0022-5193(64)90038-4

Hamilton, W. (1964b). The genetical evolution of social behaviour II. *J. Theor. Biol.* 7, 17–52. doi: 10.1016/0022-5193(64)90039-6

Harbaugh, W. T. W., Mayr, U. U., and Burghart, D. R. D. (2007). Neural responses to taxation and voluntary giving reveal motives for charitable donations. *Science* 316, 1622–1625. doi: 10.1126/science.1140738

Hart, B. L., and Hart, L. A. (1992). Reciprocal allogrooming in impala, *Aepyceros melampus*. *Anim. Behav.* 44, 1073–1083. doi: 10.1016/S0003-3472(05)80319-7

Hauert, C., and Doebeli, M. (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428, 643–646. doi: 10.1038/nature02360

Heinsohn, R., and Packer, C. (1995). Complex cooperative strategies in group-territorial African lions. *Science* 269, 1260–1262. doi: 10.1126/science.7652573

Henrich, J. (2004). Cultural group selection, coevolutionary processes and large-scale cooperation. *J. Econ. Behav. Organ.* 53, 3–35. doi: 10.1016/s0167-2681(03)00094-5

Henrich, J., and Boyd, R. (2001). Why people punish defectors: weak conformist cransmission can stabilize costly enforcement of norms in cooperative dilemmas. *J. Theor. Biol.* 208, 79–89. doi: 10.1006/jtbi.2000.2202

Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., Gintis, H. M., et al. (2001). In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev.* 91, 73–78. doi: 10.1257/aer.91.2.73

Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., Gintis, H. M., et al. (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* 28, 795–855. doi: 10.1017/s0140525x05000142

Henrich, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., et al. (2010). Markets, religion, community size and the evolution of fairness and punishment. *Science* 327, 1480–1484. doi: 10.1126/science.1182238

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, H. C., Bolyanatz, A., et al. (2006). Costly punishment across human societies. *Science* 312, 1767–1770. doi: 10.1126/science.1127333

Hill, K. R., Walker, R. S., Bozicevic, M., Eder, J., Headland, T., Hewlett, B., et al. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science* 331, 1286–1289. doi: 10.1126/science.1199071

Hoffman, E., McCabe, K., and Smith, V. (1998). Behavioral foundations of reciprocity: experimental economics and evolutionary psychology. *Econ. Inq.* 36, 335–352. doi: 10.1111/j.1465-7295.1998.tb01719.x

House, B. R., Silk, J. B., Henrich, J., Barrett, H. C., Scelza, B. A., Boyette, A. H., et al. (2013). Ontogeny of prosocial behavior across diverse societies. *Proc. Natl. Acad. Sci. U S A* 110, 14586–14591. doi: 10.1073/pnas.1221217110

Irons, W. (1998). Adaptively relevant environments versus the environment of evolutionary adaptedness. *Evol. Anthropol.* 6, 194–204. doi: 10.1002/(sici)1520-6505(1998)6:6<194::aid-evan2>3.0.co;2-b

Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., and Haselton, M. G. (2013). The evolution of error: error management, cognitive constraints and adaptive decision-making biases. *Trends Ecol. Evol.* 28, 474–481. doi: 10.1016/j.tree.2013.05.014

Johnson, D. D. P., Stopka, P., and Knights, S. (2003). Sociology: the puzzle of human cooperation. *Nature* 421, 911–912. doi: 10.1038/421911b

Johnstone, R. A., and Bshary, R. (2002). From parasitism to mutualism: partner control in asymmetric interactions. *Ecol. Lett.* 5, 634–639. doi: 10.1046/j.1461-0248.2002.00358.x

Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a constraint on profit seeking: entitlements in the market. *Am. Econ. Rev.* 76, 728–741.

Keeley, L. H. (1996). *War before Civilisation: The Myth of the Peaceful Savage.* Oxford, UK: Oxford University Press.

Kokko, H., Johnstone, R. A., and Clutton-Brock, T. H. (2001). The evolution of cooperative breeding through group augmentation. *Proc. Biol. Sci.* 268, 187–196. doi: 10.1098/rspb.2000.1349

Kollock, P. (1998). Social dilemmas: the anatomy of cooperation. *Annu. Rev. Sociol.* 24, 183–214. doi: 10.1146/annurev.soc.24.1.183

Krasnow, M. M., Delton, A. W., Tooby, J., and Cosmides, L. (2013). Meeting now suggests we will meet again: implications for debates on the evolution of cooperation. *Sci. Rep.* 3:1747. doi: 10.1038/srep01747

Kummerli, R., Burton-Chellew, M. N., Ross-Gillespie, A., and West, S. A. (2010). Resistance to extreme strategies, rather than prosocial preferences, can explain human cooperation in public goods games. *Proc. Natl. Acad. Sci. U S A* 107, 10125–10130. doi: 10.1073/pnas.1000829107

Kummerli, R., Colliard, C., Fiechter, N., Petitpierre, B., Russier, F., and Keller, L. (2007). Human cooperation in social dilemmas: comparing the snowdrift game with the prisoner's dilemma. *Proc. Biol. Sci.* 274, 2965–2970. doi: 10.1098/rspb.2007.0793

Lamba, S., and Mace, R. (2010). People recognize when they are really anonymous in an economic game. *Evol. Hum. Behav.* 31, 271–278. doi: 10.1016/j.evolhumbehav.2010.02.002

Lamba, S., and Mace, R. (2011). Demography and ecology drive variation in cooperation across human populations. *Proc. Natl. Acad. Sci. U S A* 108, 14426–14430. doi: 10.1073/pnas.1105186108

Ledyard, J. (1995). "Public goods: a survey of experimental research," in *Handbook of Experimental Economics*, eds J. Kagel and A. Roth (Princeton: Princeton University Press).

Lee, R. B. (1972). !Kung spatial organization: an ecological and historical perspective. *Hum. Ecol.* 1, 125–147. doi: 10.1007/bf01531351

Lehmann, L., Feldman, M. W., and Foster, K. R. (2008). Cultural transmission can inhibit the evolution of altruistic helping. *Am. Nat.* 172, 12–24. doi: 10.1086/587851

Lotem, A., Fishman, M., and Stone, L. (2003). From reciprocity to unconditional altruism through signalling benefits. *Proc. Biol. Sci.* 270, 199–205. doi: 10.1098/rspb.2002.2225

Luce, R. D., and Raiffa, H. (1957). *Games and Decisions.* New York: Wiley.

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

McNally, L., and Tanner, C. J. (2011). Flexible strategies, forgiveness and the evolution of generosity in one-shot encounters. *Proc. Natl. Acad. Sci. U S A* 108:E971. doi: 10.1073/pnas.1115225108

McNamara, J. M., Barta, Z., Fromhage, L., and Houston, A. I. (2008). The coevolution of choosiness and cooperation. *Nature* 451, 189–192. doi: 10.1038/nature06455

McNamara, J. M., Barta, Z., and Houston, A. I. (2004). Variation in behaviour promotes cooperation in the Prisoner's Dilemma game. *Nature* 428, 745–748. doi: 10.1038/nature02432

McNamara, J. M., and Leimar, O. (2010). Variation and the response to variation as a basis for successful cooperation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 2627–2633. doi: 10.1098/rstb.2010.0159

Milinski, M., Semmann, D., and Krambeck, H. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature* 415, 424–426. doi: 10.1038/415424a

Morgan, T. J. H., Rendell, L. E., Ehn, M., Hoppitt, W., and Laland, K. N. (2012). The evolutionary basis of human social learning. *Proc. Biol. Sci.* 279, 653–662. doi: 10.1098/rspb.2011.1172

Morin, O. (2014). Is cooperation a maladaptive by-product of social learning? The docility hypothesis reconsidered. *Biol. Theor.* 9, 286–295. doi: 10.1007/s13752-014-0181-z

Nettle, D. D., Colléony, A. A., and Cockerill, M. M. (2011). Variation in cooperative behaviour within a single city. *PLoS One* 6:e26922. doi: 10.1371/journal.pone.0026922

Noë, R., and Hammerstein, P. (1994). Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. *Behav. Ecol. Sociobiol.* 35, 1–11. doi: 10.1007/bf00167053

Nowak, M. A., and May, R. M. (1992). Evolutionary games and spatial chaos. *Nature* 359, 826–829. doi: 10.1038/359826a0

Nowak, M. A., and Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577. doi: 10.1038/31225

Oates, J., Manica, A., and Bshary, R. (2010). The shadow of the future affects cooperation in a cleaner fish. *Curr. Biol.* 20, R472–R473. doi: 10.1016/j.cub.2010.04.022

Raihani, N. J., and Bshary, R. (2011). The evolution of punishment in n-player public goods games: a volunteer's dilemma. *Evolution* 65, 2725–2728. doi: 10.1111/j.1558-5646.2011.01383.x

Raihani, N. J., and Bshary, R. (2012). A positive effect of flowers rather than eye images in a large-scale, cross-cultural dictator game. *Proc. Biol. Sci.* 279, 3556–3564. doi: 10.1098/rspb.2012.0758

Raihani, N. J., and Smith, S. (in press). Competitive helping in the real world. *Curr. Biol.*

Rand, D. G., Greene, J. D., and Nowak, M. A. (2012). Spontaneous giving and calculated greed. *Nature* 489, 427–430. doi: 10.1038/nature11467

Rand, D. G., and Nowak, M. A. (2013). Human cooperation. *Trends Cogn. Sci.* 17, 413–425. doi: 10.1016/j.tics.2013.06.003

Rand, D. G., Peysakhovich, A., Kraft-Todd, G. T., Newman, G. E., Wurzbacher, O., Nowak, M. A., et al. (2014). Social heuristics shape intuitive cooperation. *Nat. Commun.* 5:3677. doi: 10.1038/ncomms4677

Rankin, D. J., and Eggimann, F. (2009). The evolution of judgement bias in indirect reciprocity. *Proc. Biol. Sci.* 276, 1339–1345. doi: 10.1098/rspb.2008.1715

Rendell, L., Fogarty, L., Hoppitt, W. J. E., Morgan, T. J. H., Webster, M. M., and Laland, K. N. (2011). Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends Cogn. Sci.* 15, 68–76. doi: 10.1016/j.tics.2010.12.002

Richerson, P. J., and Boyd, R. (2008). *Not by Genes Alone: How Culture Transformed Human Evolution.* Chicago, IL: University of Chicago Press.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proc. Biol. Sci.* 265, 427–431. doi: 10.1098/rspb.1998.0312

Roberts, G. (2005). Cooperation through interdependence. *Anim. Behav.* 70, 901–908. doi: 10.1016/j.anbehav.2005.02.006

Roberts, G., and Sherratt, T. (1998). Development of cooperative relationships through increasing investment. *Nature* 394, 175–179. doi: 10.1038/28160

Rockenbach, B., and Milinski, M. (2011). To qualify as a social partner, humans hide severe punishment, although their observed cooperativeness is decisive. *Proc. Natl. Acad. Sci. U S A* 108, 18307–18312. doi: 10.1073/pnas.1108996108

Schroeder, K. B., Pepper, G. V., and Nettle, D. (2014). Local norms of cheating and the cultural evolution of crime and punishment: a study of two urban neighborhoods. *PeerJ* 2:e450. doi: 10.7717/peerj.450

Seielstad, M. T., Minch, E., and Cavalli-Sforza, L. L. (1998). Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* 20, 278–280. doi: 10.1038/3088

Sherratt, T. N., and Roberts, G. (1998). The evolution of generosity and choosiness in cooperative exchanges. *J. Theor. Biol.* 193, 167–177. doi: 10.1006/jtbi.1998.0703

Sherratt, T. N., Roberts, G., and Kassen, R. (2009). Evolutionary stable investment in products that confer both an individual benefit and a public good. *Front. Biosci. (Landmark Ed.)* 14, 4557–4564. doi: 10.2741/3548

Soetevent, A. R. (2005). Anonymity in giving in a natural context-a field experiment in 30 churches. *J. Pub. Econ.* 89, 2301–2323. doi: 10.1016/j.jpubeco.2004.11.002

Sommerfeld, R., Krambeck, H., and Milinski, M. (2008). Multiple gossip statements and their effect on reputation and trustworthiness. *Proc. Biol. Sci.* 275, 2529–2536. doi: 10.1098/rspb.2008.0762

Sommerfeld, R. D., Krambeck, H.-J., Semmann, D., and Milinski, M. (2007). Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl. Acad. Sci. U S A* 104, 17435–17440. doi: 10.1073/pnas.0704598104

Sylwester, K., and Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. *Biol. Lett.* 6, 659–662. doi: 10.1098/rsbl.2010.0209

Sylwester, K., and Roberts, G. (2013). Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evol. Hum. Behav.* 34, 201–206. doi: 10.1016/j.evolhumbehav.2012.11.009

Szathmáry, E., and Smith, J. M. (1995). The major evolutionary transitions. *Nature* 374, 227–232. doi: 10.1038/374227a0

Szolnoki, A., and Perc, M. (2014). Coevolutionary success-driven multigames. *EPL* 108:28004. doi: 10.1209/0295-5075/108/28004

Tooby, J., Cosmides, L., and Price, M. E. (2006). Cognitive adaptations for n-person exchange: the evolutionary roots of organizational behavior. *MDE Manage. Decis. Econ.* 27, 103–129. doi: 10.1002/mde.1287

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57. doi: 10.1086/406755

West, S. A., El Mouden, C., and Gardner, A. (2011). Sixteen common misconceptions about the evolution of cooperation in humans. *Evol. Hum. Behav.* 32, 231–262. doi: 10.1016/j.evolhumbehav.2010.08.001

Wiessner, P. (1982). "Risk, reciprocity and social influences on !Kung-San economics," in *Politics and History in Band Societies*, eds E. Leacock and R. Lee (Cambridge, UK: Cambridge University Press), 61–84.

Yamagishi, T., Terai, S., Kiyonari, T., Mifune, N., and Kanazawa, S. (2007). The social exchange heuristic: managing errors in social exchange. *Ration. Soc.* 19, 259–291. doi: 10.1177/1043463107080449

Zefferman, M. R. (2014a). Direct reciprocity under uncertainty does not explain one-shot cooperation, but demonstrates the benefits of a norm psychology. *Evol. Hum. Behav.* 35, 358–367. doi: 10.1016/j.evolhumbehav.2014.04.003

Zefferman, M. R. (2014b). When repentance and forgiveness are possible, direct reciprocity does not explain one-shot generosity, even under arbitrarily high levels of uncertainty. *Evol. Hum. Behav.* 35, 548–549. doi: 10.1016/j.evolhumbehav.2014.07.003