# Kin selection, not indirect reciprocity, explains helping those who help others in 'image scoring'

**Gilbert Roberts**

Simonside Cottage, 6 Wingates, Morpeth, NE65 8RW, UK

gilbert.roberts@yahoo.co.uk

**Abstract**

**Indirect reciprocity is often considered key to human sociality. It occurs when helping others results in an increase in reciprocal benefits from third parties. Indirect reciprocity has been proposed to operate through 'image scoring' mechanisms, whereby those who have been seen to help are more likely to receive help. Here I show that helping those who help others in image scoring models is not a mechanism of indirect reciprocity but is instead due to kin selection. Image scoring strategies are not individually adaptive because they base the decision to help upon whether a recipient has previously helped others, rather than upon whether helping would maximize their own return. Using analysis supported by evolutionary simulation, I show that helping can be favoured in image scoring systems without reciprocation, hence image scoring systems are not models of indirect reciprocity at all. Further, I use simulation of image scoring in meta-population models to illustrate how helping is a function of relatedness. I show that relatedness explains cooperation in image scoring (and associated) systems because helpers get indirect fitness benefits from helping others that share their strategy. A strategy of helping those who help others helps other copies of itself. In this way, helping behaviour ceases to be costly, and is in fact beneficial to the strategy, though not to the individual. Thus, despite some recent doubts about the usefulness of Hamilton's rule, image scoring systems demonstrate the role of relatedness in driving cooperation.**

**Keywords: indirect reciprocity; kin selection; cooperation**

## 1. Introduction

### 1.1 Background

Explaining how cooperation works outside of the confines of direct reciprocal relationships and close familial ties, especially in large-scale human societies, has become a major focus of research. Prominent amongst explanations is the notion of indirect reciprocity. This is like direct reciprocity in that it involves one individual paying some cost to benefit another, and then another paying a cost to benefit the original donor. Provided the benefits exceed the costs then both make a net profit. Where indirect reciprocity differs from direct reciprocity is in that the reciprocation comes from an individual other than the recipient (Alexander, 1987; Boyd & Richerson, 1989). The mechanism of indirect reciprocity is illustrated in Figure 1. If direct reciprocity is a system of “I scratch your back, you scratch mine” then indirect reciprocity is a case of “I scratch your back, someone else scratches mine”. This is important because indirect reciprocity is sometimes wrongly thought of as an umbrella term for fitness benefits arising through third parties.

A key advance came with the proposal of a mechanism by which indirect reciprocity might work: if individuals who help others acquire a cooperative reputation, or ‘image score’ (Nowak & Sigmund, 1998); and if potential donors use these image scores to decide whether to help; then simulations show that this can lead to mutually beneficial cooperation (Nowak & Sigmund, 2005). Crucially, and in contrast to direct reciprocity (Axelrod & Hamilton, 1981), this is said to occur even when donors never meet their recipients again to get a direct return (Nowak & Sigmund, 1998). This appears to be supported by experimental evidence that people give preferentially to those with positive image scores (Milinski, 2016; Milinski, Semmann, Bakker, & Krambeck, 2001; Seinen & Schram, 2006; Semmann, Krambeck, & Milinski, 2004; Wedekind & Milinski, 2000). Indirect reciprocity has consequently become established as a “decisive” (Nowak & Sigmund, 1998) explanation for

cooperation, especially for human sociality, accounting for large-scale fluid societies; for why individuals care about being seen to help others; for language; and for morality (Nowak & Sigmund, 2005). Despite being fundamentally criticized (Leimar & Hammerstein, 2001), the literature on indirect reciprocity has burgeoned (e.g. (Brandt & Sigmund, 2005; Ohtsuki & Iwasa, 2005, 2006; Pacheco, Santos, & Chalub, 2006; Panchanathan & Boyd, 2004; Suzuki & Akiyama, 2008; Takahashi & Mashima, 2006; Tanimoto, 2007; Uchida & Sigmund, 2010)).

Here, I show that what has been called indirect reciprocity is instead driven by relatedness. That is, individuals do not help other helpers to receive something back from third parties; instead, they help because this furthers the spread of a strategy of helping those who help others (see Figure 1 for a comparison of the mechanisms of indirect reciprocity, image scoring and relatedness). I demonstrate this claim by showing first how 'image scoring' models and experiments do not provide a mechanism for the evolution of indirect reciprocity; secondly showing that what has been termed 'indirect reciprocity' is not driven by reciprocity at all; and thirdly showing the key role of relatedness in driving cooperation in image scoring (and 'standing') systems. In each case I draw on game theoretic analyses, evolutionary simulation, and empirical data.

I begin with the dilemma that image scoring is not an adaptive strategy. For cooperation (as referred to in game theory as a behaviour which benefits others but which may involve a cost to the actor (West, Griffin, & Gardner, 2007)) to bring fitness benefits, there must be a relationship between giving and receiving such that the costs of cooperating are outweighed by the benefits of receiving, relative to non-cooperators. Image scoring offers a mechanism for linking donations with returns through being able to identify and direct cooperation towards those who have cooperated. Image scorers are discriminating: they cooperate according to the recipient image score, which is a function of how cooperative the recipient

itself has been. However, the logic of image scoring contains an in-built inconsistency which means it cannot explain indirect reciprocity.

This inconsistency centres on the fact that image scoring strategies do not evolve when individuals are self-interested, i.e. when they are maximizing their payoff. This can readily be shown analytically. A population of image scorers (i.e. those that take into account the image score of a recipient in deciding whether to give) is invaded by those who help others indiscriminately (Leimar & Hammerstein, 2001). This is because when a population is dominated by individuals playing a strategy of 'help those with a positive image score', an individual will receive more when they take more opportunities to cooperate (Leimar & Hammerstein, 2001). It pays to have the maximum image score since this attracts the largest number of donations. Crucially this is true regardless of the image score of the recipient.  Thus, it will pay individuals to help on every move, even if this means helping defectors rather than playing the discriminating image scoring strategy. Hence, image scoring is not in an individuals' best interests (Leimar & Hammerstein, 2001).

If image scoring were adaptive then agent-based evolutionary simulation models should provide evidence that IS strategies invade and reach a stable equilibrium in which they are at fixation in the population. The simulations of Nowak and Sigmund (Nowak & Sigmund, 1998) actually show unstable behaviour, and their methods have been questioned (Leimar & Hammerstein, 2001). Following Leimar and Hammerstein's island-based method (Leimar & Hammerstein, 2001), and using a simplified system in which image scores are 0 and 1 depending on the last move, Roberts (2015) confirmed that the dynamics of IS are very unstable (Ref 22, Fig 3a). Further, the strategies which evolved were not discriminating, as required if they follow image scoring rules. Rather they showed virtually no preference for giving to those who had previously cooperated or defected (Ref 22, Fig 4): with the parameters used in this simulation, the mean image score of those donated to was -0.13 ±

0.03. A mean recipient image score so close to zero implies that potential donors playing IS are showing no discrimination by image score. Without discrimination, cooperation cannot be sustained because there is no correlation between giving and receiving (Gilbert Roberts, 2015).

Empirical evidence supporting image scoring as an adaptive strategy is weaker than often supposed. There appears to be widespread evidence that people behave more prosocially when they are observed, although this may be more nuanced than commonly thought (Bradley, Lawrence, & Ferguson, 2018). However, this cannot be taken as evidence for indirect reciprocity as opposed to the range of other possible benefits arising from being observed, such as reputation-based partner choice (Roberts, 1998; Sylwester & Roberts, 2010, 2013). Students in experimental economic games do show at least partial discrimination, tending to give more often to those with high image scores (Wedekind & Milinski, 2000). However, while this has been interpreted as evidence for indirect reciprocity via image scoring, the experiments do not necessarily show that those who donate more get higher payoffs. Groups with more donors can do better (Milinski, Semmann, & Krambeck, 2002; Wedekind & Milinski, 2000), but this has to be true because the benefit of receiving a donation is always set by the experimenter to be greater than the cost of donating (otherwise reciprocity would bring no advantage), hence the relationship between donations and group profit is an input rather than an interesting result of the experiment. Indirect reciprocity requires that individuals do better than others within their own groups by engaging in donations, but this is not always the case. For example, one leading paper reporting image scoring does not find evidence for image scoring giving higher payoffs than other strategies. A benefit to individuals has been shown only when indirect reciprocity is combined with another game (Milinski et al., 2002; Wedekind & Braithwaite, 2002; Wedekind & Milinski, 2000). Combining games in this way allows individuals to make a profit overall despite

making a loss in the social dilemma part (Roberts, 1998). No clear evidence has been provided that playing an image scoring strategy leads to net benefits via indirect reciprocity.

Furthermore, even if any of the studies purporting to demonstrate image scoring did show that those who gave more did better, this would not be evidence for image scoring: it could be evidence of playing the indiscriminate strategy referred to in the previous sections. What we really need to show is not that those who give most get most, but that those who discriminate most have a higher payoff than those who give to anyone. No paper presents an analysis of this, so no empirical demonstration of indirect reciprocity exists.

These approaches together show that image scoring is against individual strategic interests; it leads to a conflict of interest between individual and strategy; and it cannot explain indirect reciprocity.  Theorists have developed strategies that can outperform image scoring, including ‘standing’ strategies (Leimar & Hammerstein, 2001; Sugden, 1986) and the ‘leading eight’ set of strategies (Ohtsuki & Iwasa, 2006). Such strategies take a more sophisticated view of defections so that agents do not lose their good standing upon ‘justified’ defection (Panchanathan & Boyd, 2003). In this paper, I focus on image scoring for several reasons. First, the classic models of image scoring (Nowak & Sigmund, 1998) set in train a large theoretical and empirical field (Nowak & Sigmund, 2005), so if we are to understand that field, we must first understand its foundations. Secondly, despite many developments, the image scoring strategy continues to be by far the most heavily cited (e.g. (Hilbe, Schmid, Tkadlec, Chatterjee, & Nowak, 2018; Whitaker, Colombo, & Rand, 2018)), so those continuing to cite image scoring need to understand it. Thirdly, there remains a focus amongst empiricists on trying to find evidence for image scoring rather than more sophisticated strategies (e.g. Milinski et al., 2001). On this basis, I use analysis and evolutionary simulation to test the hypotheses (1) that image scoring is not a form of reciprocity; and (2) that helping in these systems is actually a function of relatedness.

### 1.2 Hypothesis 1: Helping in image scoring systems is not a form of reciprocity

If cooperation in image scoring systems is due to indirect reciprocity, then perhaps the most obvious criterion would be that there must be evidence that reciprocation is driving the benefits of cooperation: i.e. that strategically, individuals donate to profit from receiving more (where b >c). If donation still occurred in the absence of reciprocation then indirect reciprocity could not be the explanation.

The basis of any form of reciprocity is that one play involves a cost and that this only results in a net profit when another individual plays and returns the act of help. We can ask whether image scoring is a strategy of (indirect) reciprocity or whether it involves a different selective mechanism. Helpers always pay the cost of helping but with a certain probability they are also helping other helpers. This means that while paying a cost they also benefit their own strategy. So given a probability that they help individuals carrying copies of the same strategy, the strategy can make a net profit through helping. This ability for helping to pay off in one shot interactions is unique to the strategy of helping those who help others. Figure 2 illustrates this with the case of an individual who helps other helpers meeting another individual with the same strategy.

Image scoring is distinct from reciprocity because it requires no reciprocation to work. This is because when an individual carrying a ‘help those who helps’ strategy helps another individual carrying the same strategy, then it has a stake in the benefits, regardless of whether there is any reciprocation. Helping can be favoured when the donor is sufficiently closely related to the recipient. This probability has been referred to as ‘social acquaintanceship’ (Rand & Nowak, 2013) (effectively the chance of an agent who helps others meeting another agent who helps others) but can also be described as relatedness.

Crucially, evolutionary simulations use the payoff accrued to strategies to determine how many individuals in next generation carry that strategy: this is how strategies evolve and how we find which strategies are evolutionarily stable. Individuals playing ‘image scoring’ by helping those who help others are actually helping copies of their own strategy as carried by their recipients, and so image scoring can be seen to be a form of relatedness in which the image scores are markers showing that individuals share the ‘help those who help others’ strategy. Reciprocity is unnecessary for a “help those who help others” system because helping brings benefits to the helping strategy itself: every helping act actually provides a net benefit to the helping strategy when both actor and recipient are carriers.

Here, I use evolutionary simulation to test the robustness of this inference. I ask whether agents will help even when they cannot get a return through direct or indirect reciprocation. In all previous considerations of image scoring, agents do not have information about whether it is the last move. In reciprocity, this is predicted to be fatal since no agent should cooperate on a last move (which becomes equivalent in principle to a one-shot game). However, from the analysis above, it follows that image scoring strategies of helping those who help others are predicted to help on the last move when doing so increases the payoff of the strategy, at the expense of the helping individual.

### 1.3 Hypothesis 2: Helping in image scoring systems is a function of relatedness

My second hypothesis is that helping in image scoring systems is actually explained by relatedness. Individuals help other helpers, not in the expectation of reciprocal help, but because by doing so they are helping copies of the “help those who help others” strategy. Thus, it is a mistake to look for direct fitness benefits through reciprocation. Instead individuals get indirect fitness benefits. Indirect fitness benefits accrue when an individual altruist is related to a recipient. If helping behaviour is a function of relatedness rather than

reciprocity, then donations should increase as average relatedness increases. One way to vary relatedness in models is to vary the demographic structure. Here I do this by varying the proportion of local versus global contributions to reproductive success. This technique has the advantage of placing our analyses in the context of the critique by Leimar and Hammerstein (Leimar & Hammerstein, 2001). We can predict that island structure makes it harder for relatedness to operate, because helping other island members has a lower effect on the frequency of a strategy in the next generation. Whereas in a single population, helping another individual of the 'help those who help others' strategy means the benefit goes entirely to the reproductive success of the actor's (and recipient's) strategy, if the next generation is formed of only a portion of a local population, then the relatedness of a recipient will effectively be reduced. Thus, as island effects increase, so it becomes harder to satisfy the proposed condition $rb > c$ for helping to occur.

## 2. Methods

Simulation methods were adapted from those in reference (Gilbert Roberts, 2015). Pseudocode for the simulations is given in Supplementary Information. I modelled indirect reciprocity using a form of first-order reputation indexing known as image scoring (Nowak & Sigmund, 1998). Scores were initialised at 0 for each individual in each generation reflecting the fact that individuals begin their lives without experience. Scores then became -1 following defection and 1 following cooperation. Image scores depended only upon the last move. Strategies were defined by their thresholds k, where k = -1 denotes individuals that cooperated unconditionally; k = 0 strategists cooperated provided their partner had never played or had cooperated; k = 1 strategists cooperated provided the partner had cooperated; and k = 2 strategists never cooperated. Individuals interacted in groups of size g with default g = 100.

Interactions proceeded by the program randomly choosing a focal individual (i.e. an individual upon whom operations are performed by the program) which then had an opportunity to donate to a partner from the same group, conditional upon its own strategy and the partner's image score. Focals were assigned partners at random from their group by the program. This interaction process was repeated such that every group member had on average *m* meetings with every other, either as donor or recipient, where by default m = 1. If the focal donated, it incurred cost c = -1 and the partner received benefit b = 2. The donation cost-benefit ratio was set to 1:2. To avoid negative payoffs (which should not occur in reality provided cooperating costs a small proportion of available resources), all focals had their payoffs incremented by 1 each time they played.

A meta-population structure was used in which p = 10,000 individuals were distributed evenly over i = 100 islands, where islands are units in which reproduction is structured. Note that the default island size was equal to one group, but the parameters were distinct so that interaction group size could be varied without affecting reproduction and drift. Simulations started with all strategies at *k*=2 (always defect) and neutral image scores. Simulations were evolutionary in that those strategies accumulating the highest payoff produced most offspring. Reproduction was based on the success of a strategy both within and between islands. The payoffs of all strategies were summed, then individuals were assigned strategies in the next discrete generation in proportion to the relative payoff of each strategy. An individual for the next generation was derived (by default) locally with probability 0.9 and globally with probability 0.1. Population size was therefore kept constant, with each generation being fully replaced by the next. Reproduction was accompanied by mutation: with probability $\mu$ = 0.01 an individual's strategy was replaced at random (any strategy could change to any other strategy without having to track through an imposed sequence). Simulations were run for 10,000 generations.

To test Hypothesis 1, I used the above methods to simulate a helping game in which agents had the flexibility to express a different behaviour on their last move. The simulation method was modified to produce two stages. In the new second stage, agents were selected randomly as in stage 1, but after playing one move (in which they either cooperated or defected, depending on their strategy), they were marked so that they could not be selected as a recipient for any other agent. Behaviour in stage 2 was coded separately, so an agent could help in stage 1, be assigned a positive image score, but then decline to help (defect) in stage 2 when there was no chance of receiving anything afterwards. If helping is selected for because image scores encourage reciprocity, then agents should help in stage 1 but not stage 2.

To test Hypothesis 2, I used the same simulation methods but varied the proportion of local versus global contributions to reproductive success. That is, during the reproduction procedure, an individual for the next generation was derived by default locally with probability 0.9 and globally with probability 0.1 (i.e. the next generation reflected strongly the relative reproductive successes of strategies within the local rather than the global population), but in this set of simulations, these probabilities were varied so as to vary the extent to which individuals could help their relatives.

## 3. Results

### 3.1 Hypothesis 1

Analyzing the results of the simulations I found that the dynamics fit the well-reported pattern of cooperation establishing through conditional strategies (k=0 and k=1) then unconditional cooperation leaking into the conditionally cooperative environment followed by the invasion of all defection, and so the cycle continued (e.g. Sherratt & Roberts, 2001). I found no difference in the percentage composition of strategies between those expressed during the

body of the simulation and those expressed on the last move when an agent could not receive anything after donating (Figure 3). Cooperation rates were 73.36±1.93 and 70.16±1.71 respectively in the body versus the last move of the simulations (means and standard errors across 10 simulations, with means calculated across all 10000 generations in each) Thus, having no chance of receiving does not provide a selection pressure to avoid helping. This is in contrast to the prediction from reciprocity theory which is by definition based on there being a return.

### 3.2 Hypothesis 2

As predicted, I found that cooperation declines as relatedness declines. This can be seen in Figure 4 where the proportion of helping behaviour in relation to the proportion that a local population contributes to the next generation.

## 4 Discussion

Previous analytical, simulation and experimental evidence indicates that image scoring strategies are inconsistent with individual adaptation. On the basis of this, I hypothesized that image scoring strategies are not mechanisms of indirect reciprocity at all and are actually a function of relatedness. Two lines of evidence from the evolutionary simulations in this paper support this hypothesis.

First, in the simulations where agents could change strategy on the basis of it being their last move, they continued to donate. If it is the last move (or a one-shot game) then giving cannot be explained by reciprocity. A good example of a one-shot game is by van Apeldoorn & Schram (2016). Here people appear to be rewarding those with high image scores (hence it is interpreted as evidence of image scoring in the paper). However, there is no prospect of receiving a return and so no explanation as to why individuals might do this.

There is no mechanism for those that discriminate to make a payback, so indirect reciprocity cannot explain the discriminating behaviour in terms of individual benefit.

Secondly, varying the ‘island’ structure of the simulations showed that image scoring was only found when helping other agents of the same strategy type resulted in the spread of that strategy through time. In Leimar and Hammerstein’s (Leimar & Hammerstein, 2001) original investigation of image scoring in the context of an island population, they proposed that drift was responsible for Nowak and Sigmund’s (Nowak & Sigmund, 1998) finding of cooperation. By introducing population structure they argued that image scoring could not give rise to cooperation without drift. Here I propose instead that the causal factor is relatedness: when this is reduced (by reducing the proportion of offspring that result from success within an island), so cooperation declines.

To summarize, what has been termed ‘indirect reciprocity’ in a large literature cannot be explained by individual adaptation; does not explain why help is directed discriminatingly; is not even a form of reciprocity; and is actually a function of relatedness. It works not because of getting a reciprocal benefit but by helping copies of the strategy of “help those who help others”.

Relatedness provides the only adaptive explanation for existing empirical results. In economic game experiments, selection on individual fitness cannot explain why individuals donate, because they do not do better by doing so (Wedekind & Milinski, 2000). I propose that individuals are helping others who carry the same “help those who help” strategy. As such, they are helping relatives, and this is costly for the individual yet beneficial to the strategy.

If ‘image scoring’ systems are actually driven by relatedness and not reciprocity then the condition for helping to occur must include a term for relatedness. By Hamilton’s rule altruism can be favoured when

$$rb > c$$

namely when the costs *c* to the actor are more than offset by the benefits *b* to a recipient scaled by their relatedness *r* (Hamilton, 1964). The commonest way in which altruism evolves is through genealogical relatedness. However, relatedness quantifies the genetic similarity between any two individuals (relative to the average similarity of all individuals in the population (West et al., 2007)) and is not confined to cases of close genealogical descent. Hamilton’s rule takes the same form as the condition for cooperation in image scoring systems derived by Nowak and Sigmund, namely:

$$qb > c$$

where *q* is ‘social acquaintanceship’ (or the probability of knowing the image score of a recipient) (Nowak, 2006). I argue that “social acquaintanceship” is actually relatedness. Individuals should pay the cost of helping when this is offset by the benefit to the recipient multiplied by their chance of sharing the same strategy. This is selected for through indirect fitness benefits, and not through indirect reciprocity.

It has long been recognised that the ‘standing’ strategy, in which individuals help those who help other helpers (Sugden, 1986)) may have superior properties to image scoring, at least in theory (Leimar & Hammerstein, 2001), if not in practice (Milinski et al., 2001). This strategy is in effect a more discriminating version of image scoring: if image scorers help all those who have helped others, those following the standing strategy help a subset of these, i.e. those who have helped other helpers. Thus, standing also works through relatedness, with *r* being higher than for image scorers. An easy way to see this is to consider a standing strategist on their final play with no chance of receiving a return: they should nevertheless

donate to another individual who helps helpers when by doing so this furthers their own strategy. Similarly, selecting a recipient from a range of possibilities has been shown to allow for greater discrimination (Roberts, 2015), but again this is actually a function of relatedness whereby donors experience higher r by avoiding those that do not carry the same strategy.

Recognising this resolves the fundamental problem of how cooperation in image scoring systems is individually costly yet beneficial to the strategy: donating is only beneficial in inclusive fitness terms. Reciprocity is an illusion resulting from the fact that individuals receive after revealing their strategy by helping. ‘Image scores’ are not reputation indices but are actually kinship cues. Where genealogical relatedness is not found, they may be ‘green beards’: they display a distinguishing phenotype (helping others) allowing individuals to direct help to those who are genetically related (Dawkins, 1976; West & Gardner, 2010). Green beard altruism is expected to be rare, because individuals might ‘cheat’, displaying a green beard when they do not have the altruistic gene (but see Keller & Ross, 1998). However, as recognised by Dawkins himself, helping others is a uniquely honest green beard in that the expression of the helping behaviour is the green beard itself. So helping those who help others is a special case that can promote the spread of altruistic behaviour through the green beard effect of recognizing and favouring those that share the strategy, even when they are not genealogical kin. Recent evidence suggests that hunter gatherers, for example, have social networks that are based as much on similarity in helping behaviour as they are on genealogical kinship (Apicella, Marlowe, Fowler, & Christakis, 2012).

Several authors have recognized that if individuals are helping those displaying the green beard of helping, then this may lead to reciprocal help (Biernaskie, Grafen, & Perry, 2014; Dugatkin, Wilson, Farrand, & Wilkens, 1994; Nee, 1989; Price, 2006; Trivers, 2006). This may well be the case, but this paper is actually saying that what has been studied by those

using image scoring based models and experiments is not reciprocity at all. Hence the connections between reciprocity theory and green beard theory represent a separate strand of inquiry.

We can see that ‘image scores’ simply provide ways of discriminating other carriers of the strategy of helping others (relatives). This is important because indirect reciprocity theory has been employed in proposing how individuals make moral judgements. When individuals give, they are thought to use a potential recipient’s image score to decide whether they are ‘worthy’ of receiving. This approach has gone so far as to “define goodness” (Ohtsuki & Iwasa, 2005) and to identify the “leading eight” ‘social norms’ of 4096 possibilities identified (Ohtsuki & Iwasa, 2006). When individuals fail to cooperate, these defections have been described as ‘justified’ and ‘unjustified’ on the basis of whether the partner was cooperative or not (Panchanathan & Boyd, 2003). The results of this paper demonstrate that these kinds of analyses are mistaken in their quest for a theoretical basis of morality within the context of who is helped: in fact, they are analyses of when indirect fitness supports cooperation. This is a fundamentally different mechanism and explanation.

I have considered image scoring in this paper, but strategies that better identify relatives will do better, for example I predict that the success of the ‘stern judging’ strategy (Pacheco et al., 2006) is not to do with morality, reputations and norms (as presented in (Santos, Santos, & Pacheco, 2018)) but is because it is particularly good at directing help to relatives. Giving to non-relatives reduces the effectiveness of the strategy, but according to Hamilton’s Rule (Hamilton, 1964), strategies of helping relatives will still spread when there are inclusive fitness benefits. The problem with image scoring is entirely different. It is that image scoring is intended as a strategy of indirect reciprocity, but is actually costly to the individual, and is

outcompeted by other strategies in indirect reciprocity games (Leimar & Hammerstein, 2001).

Indirect reciprocity has become synonymous with a concern for reputation (Brandt & Sigmund, 2005; Nowak & Sigmund, 2005), yet it is unlikely that indirect reciprocity has any significant role in explaining displays of generosity (e.g. charitable donations). Evidence is much weaker than is typically thought, partly because some authors reference ‘image scoring’ or ‘indirect reciprocity’ when they actually just mean increases in pro-social behaviour in a social environment (e.g. Bshary & Grutter, 2006; Yoeli, Hoffman, Rand, & Nowak, 2013), despite the fact that such phenomena are simply elements of a widespread ‘observability effect’ (Bradley et al., 2018; Burnham & Hare, 2007). Furthermore, as discussed here, evidence of helping those who have helped others in economic games is not necessarily support for image scoring or indirect reciprocity as has been claimed (Wedekind & Milinski, 2000) because such evidence is equally compatible with kinship. Some experiments use combinations of games of different types (Milinski et al., 2002), but these cannot be taken as evidence for indirect reciprocity because costs borne in indirect reciprocity games are not recouped in indirect reciprocity games. This takes us back to Figure 1, which shows how indirect reciprocity is a simple modification of direct reciprocity in which the reciprocator are different individuals.

The claim that indirect reciprocity can work in large fluid societies and so support cooperation without individuals even meeting again was a paradigm-changing result (Nowak & Sigmund, 1998)(*3*). This stood in stark contrast to the dominant role of repeated interactions in the evolution of cooperation (Axelrod & Hamilton, 1981). Given the evidence presented here that donations are explained by relatedness, this claim turns out to be false. If indirect reciprocity is a significant explanation for cooperation at all then it’s role may be

better seen in the original ideas of Alexander (1987) and the reputation models of Pollock and Dugatkin (1992) where observational information is used to inform play on first meeting but crucially the benefits come through finding a partner for direct reciprocity, and not through indirect reciprocity. Whether the fitness benefits of indirect reciprocation can support widespread cooperation must be radically revisited, and we may need to return to the original models of indirect reciprocity which showed it was not adaptive (Boyd & Richerson, 1989).

To conclude, although the role of inclusive fitness in sociality has recently been challenged (Nowak, Tarnita, & Wilson, 2010), 'image scoring' systems may turn out to be one of the most elegant demonstrations of the causal influence of relatedness in social behaviour. This is significant because few new ideas have been as influential in evolutionary biology in recent decades. Given that image scoring has been cited as a decisive factor in the evolution of human large-scale sociality, including morality (who is a worthy recipient) and language (how we share such information) (Alexander, 1987; Nowak & Sigmund, 2005; Santos et al., 2018) these issues must be revisited. It also has fascinating implications for the role of cultural evolution, whereby strategies of helping helpers may be passed on within generations by social learning (Boyd & Richerson, 1989). Even when these strategies are individually costly, strategies of helping other helpers have the special property of being able to spread by helping copies of themselves. This offers a potential explanation for prosocial norms including generosity and fairness.

Funding: This This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

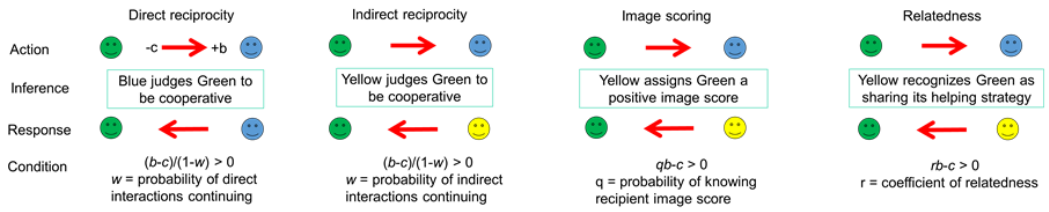


**Figure 1. Reciprocity and relatedness.**

A comparison of the processes involved in direct reciprocity, indirect reciprocity, image scoring and relatedness, and of the conditions in which they can account for cooperation. Arrows show donations. When an individual performs a donation ('cooperates' or plays 'C'), it pays a cost *c* and provides a benefit *b* to another individual. Direct reciprocity brings a direct fitness benefit when partners respond by cooperating and when the probability of interactions continuing *w* is sufficiently large. Indirect reciprocity is expected to have the corresponding condition, but where returns come from individuals other than the recipient of the actor's donation. Image scoring (Nowak & Sigmund, 1998) is a postulated mechanism of indirect reciprocity which brings a direct fitness benefit when the probability of knowing the image score of a recipient (*q* or 'social acquaintanceship' (Rand & Nowak, 2013)) is sufficiently large (Nowak & Sigmund, 1998). Finally, where strategies help copies of the same strategy, cooperating brings an indirect fitness benefit according to Hamilton's rule. Note the similar form of these conditions.

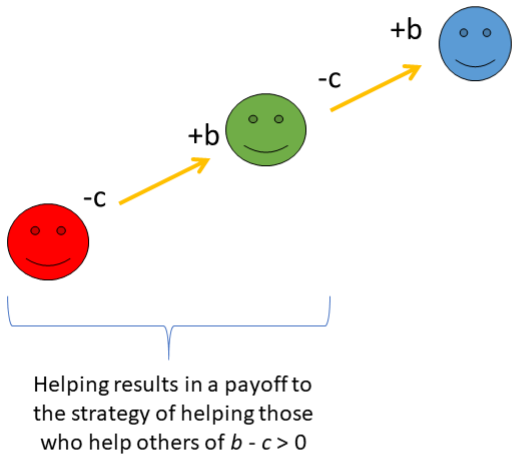


**Figure 2. Helping in 'image scoring' models can occur without reciprocity**

Consider here that individual Red observes Green to have helped Blue (where arrows represent helping). Green then has a positive image score and Red, playing an image scoring strategy, pays cost *c* to benefit Green by amount *b* (where it is assumed that b>c for reciprocity to pay*)*. Assuming costs and benefits translate into fitness, this means the fitness of Red is reduced by *c* when it helps. However, the payoff to the strategy of helping those who help others is actually increased on every helping act by amount *b* - *c*. This behavior could occur without reciprocity, direct or indirect, when Red and Green are related, according to Hamilton's Rule, *rb-c* >0 (Hamilton, 1964). That is, Red pays a cost but Red also gets a benefit in proportion to the probability that Green shares the "help those who help others" strategy.

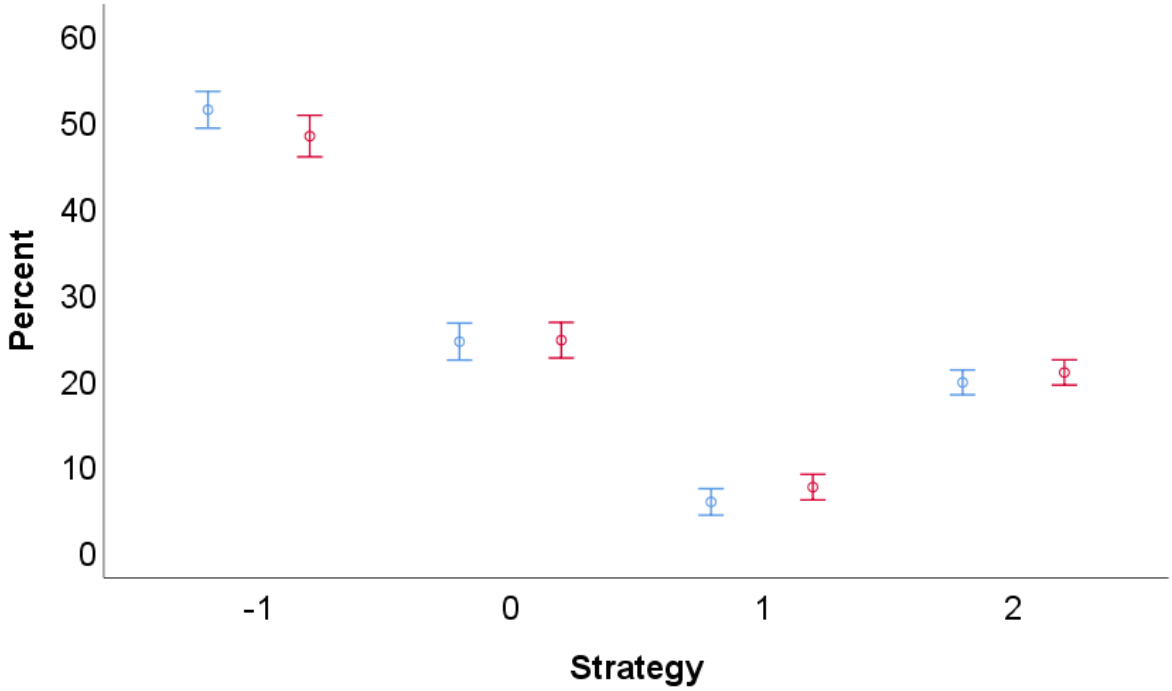


**Figure 3. Helping continues without the possibility for reciprocity**

Plotted are means and standard errors of the percentage composition of a simulated population by strategies k=-1, 0, 1 and 2. Percentages were calculated over 10,000 generations for each of 10 simulations. Strategies expressed in the body of the simulation are plotted in blue, strategies as expressed in the last move, where reciprocation was impossible, are plotted in red. For simulation methods and parameter values see the Methods.

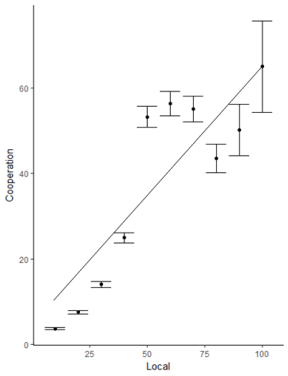


**Figure 4. Population structure, relatedness and cooperation.**

The percentage of helping behaviour in relation to the percentage that a local population (versus the global population) contributes to the next generation. Percent help is expressed as the mean (and standard error) across 20 simulations of the proportion of moves that were cooperative in generation 10,000. Simulations were run with a population of individuals divided into 10 islands with N=100 on each. Reproduction was varied so that strategy frequencies in the next generation reflected strategy payoffs on the local island and those averaged across all islands. Thus, x% was derived from the local island while 1-x% took their strategy from the global population. Relatedness between donors and receivers increases as the reproduction regime favours local proportions because this means that helping others that carry the same strategy will have more impact on the success of that strategy. The line shows the level of cooperation predicted from the change in relatedness as local contributions change. As predicted, cooperation declines as local contributions, and thereby relatedness, declines. There appears to be a threshold at around a 50:50 ratio of local to global.

## References


Alexander, R. D. (1987). *The Biology of Moral Systems*. New York: Aldine de Gruyter.

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers. *Nature, 481*, 497.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science, 211*, 1390-1396.

Biernaskie, J. M., Grafen, A., & Perry, J. C. (2014). *The evolution of index signals to avoid the cost of dishonesty* (Vol. 281).

Boyd, R., & Richerson, P. J. (1989). The Evolution of Indirect Reciprocity. *Social Networks, 11*(3), 213-236.

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society B: Biological Sciences, 285*(1875). doi:10.1098/rspb.2018.0116

Brandt, H., & Sigmund, K. (2005). Indirect reciprocity, image scoring, and moral hazard. *Proceedings of the National Academy of Sciences of the United States of America, 102*(7), 2666-2670.

Bshary, R., & Grutter, A. S. (2006). Image scoring and cooperation in a cleaner fish mutualism. *Nature, 441*(7096), 975-978.

Burnham, T. C., & Hare, B. (2007). Engineering human cooperation - Does involuntary neural activation increase public goods contributions? *Human Nature-an Interdisciplinary Biosocial Perspective, 18*(2), 88-108. doi:10.1007/s12110-007-9012-2

Dawkins, R. (1976). *The Selfish Gene*.

Dugatkin, L. A., Wilson, D. S., Farrand, L., & Wilkens, R. T. (1994). Altruism, tit for tat and 'outlaw' genes. *Evolutionary Ecology, 8*(4), 431-437. doi:10.1007/BF01238193

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology, 7*, 1-52.

Hilbe, C., Schmid, L., Tkadlec, J., Chatterjee, K., & Nowak, M. A. (2018). Indirect reciprocity with private, noisy, and incomplete information. *Proceedings of the National Academy of Sciences, 115*(48), 12241-12246. doi:10.1073/pnas.1810565115

Keller, L., & Ross, K. G. (1998). Selfish genes: a green beard in the red fire ant. *Nature, 394*(6693), 573-575.

Leimar, O., & Hammerstein, P. (2001). Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*(1468), 745-753.

Milinski, M. (2016). Reputation, a universal currency for human social interactions. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 371*(1687). doi:10.1098/rstb.2015.0100

Milinski, M., Semmann, D., Bakker, T. C. M., & Krambeck, H. J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society of London Series B-Biological Sciences, 268*(1484), 2495-2501.

Milinski, M., Semmann, D., & Krambeck, H. J. (2002). Reputation helps solve the 'tragedy of the commons'. *Nature, 415*(6870), 424-426.

Nee, S. (1989). Does Hamilton's rule describe the evolution of reciprocal altruism? *Journal of Theoretical Biology, 141*(1), 81-91. doi:http://dx.doi.org/10.1016/S0022-5193(89)80010-4

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314*(5805), 1560-1563.

Nowak, M. A., & Sigmund, K. (1998). Evolution of indirect reciprocity by image scoring. *Nature, 393*(6685), 573-577.

Nowak, M. A., & Sigmund, K. (2005). Evolution of indirect reciprocity. *Nature, 437*, 1291-1297.

Nowak, M. A., Tarnita, C. E., & Wilson, E. O. (2010). The evolution of eusociality. *Nature, 466*(7310), 1057-1062. doi:http://www.nature.com/nature/journal/v466/n7310/abs/nature09205.html#supplementary-information

Ohtsuki, H., & Iwasa, Y. (2005). How should we define goodness? reputation dynamics in indirect reciprocity (vol 231, pg 107, 2004). *Journal of Theoretical Biology, 232*(4), 451-451.

Ohtsuki, H., & Iwasa, Y. (2006). The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology, 239*(4), 435-444.

Pacheco, J. M., Santos, F. C., & Chalub, F. (2006). Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *Plos Computational Biology, 2*(12), 1634-1638. doi:e178

Panchanathan, K., & Boyd, R. (2003). A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology, 224*(1), 115-126.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432*(7016), 499-502.

Pollock, G. B., & Dugatkin, L. A. (1992). Reciprocity and the evolution of reputation. *Journal of Theoretical Biology, 159*, 25-37.

Price, M. E. (2006). Monitoring, reputation, and ‘greenbeard’ reciprocity in a Shuar work team. *Journal of Organizational Behavior, 27*(2), 201-219. doi:10.1002/job.347

Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences, 17*(8), 413-425.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London Series B-Biological Sciences, 265*(1394), 427-431.

Roberts, G. (2015). Partner Choice Drives the Evolution of Cooperation via Indirect Reciprocity. *PLoS ONE, 10*(6), e0129442. doi:10.1371/journal.pone.0129442

Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature, 555*, 242.

Seinen, I., & Schram, A. (2006). Social status and group norms: Indirect reciprocity in a repeated helping experiment. *European Economic Review, 50*(3), 581-602.

Semmann, D., Krambeck, H. J., & Milinski, M. (2004). Strategic investment in reputation. *Behavioral Ecology and Sociobiology, 56*(3), 248-252.

Sherratt, T. N., & Roberts, G. (2001). The role of phenotypic defectors in stabilizing reciprocal altruism. *Behavioral Ecology, 12*, 313-317.

Sugden, R. (1986). *The Economics of Rights, Cooperation and Welfare*. Oxford: Blackwell.

Suzuki, S., & Akiyama, E. (2008). Evolutionary stability of first-order-information indirect reciprocity in sizable groups. *Theoretical Population Biology, 73*(3), 426-436.

Sylwester, K., & Roberts, G. (2010). Cooperators benefit through reputation-based partner choice in economic games. *Biology Letters, 6*(5), 659-662. doi:10.1098/rsbl.2010.0209

Sylwester, K., & Roberts, G. (2013). Reputation-based partner choice is an effective alternative to indirect reciprocity in solving social dilemmas. *Evolution and Human Behavior, 34*(3), 201-206. doi:http://dx.doi.org/10.1016/j.evolhumbehav.2012.11.009

Takahashi, N., & Mashima, R. (2006). The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity. *Journal of Theoretical Biology, 243*(3), 418-436.

Tanimoto, J. (2007). A study of indirect reciprocity involving a reputation system or a simple tag system in a one-shot, multi-player game. *Biosystems, 90*(3), 856-869.

Trivers, R. (2006). Reciprocal altruism: 30 years later. In P. M. Kappeler & C. P. van Schaik (Eds.), *Cooperation in Primates and Humans: Mechanisms and Evolution* (pp. 67-83). Berlin, Heidelberg: Springer Berlin Heidelberg.

Uchida, S., & Sigmund, K. (2010). The competition of assessment rules for indirect reciprocity. *Journal of Theoretical Biology, 263*(1), 13-19. doi:DOI: 10.1016/j.jtbi.2009.11.013

van Apeldoorn, J., & Schram, A. (2016). Indirect Reciprocity; A Field Experiment. *PLOS ONE, 11*(4), e0152076. doi:10.1371/journal.pone.0152076

Wedekind, C., & Braithwaite, V. A. (2002). The Long-Term Benefits of Human Generosity in Indirect Reciprocity. *Current biology : CB, 12*(12), 1012-1015.

Wedekind, C., & Milinski, M. (2000). Cooperation through image scoring in humans. *Science, 288*(5467), 850-852.

West, S. A., & Gardner, A. (2010). Altruism, spite, and greenbeards. *Science, 327*(5971), 1341-1344. doi:10.1126/science.1178332

10.1126/science.1178332.</p>

West, S. A., Griffin, A. S., & Gardner, A. (2007). Social semantics: altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology, 20*(2), 415-432. doi:10.1111/j.1420-9101.2006.01258.x

Whitaker, R. M., Colombo, G. B., & Rand, D. G. (2018). Indirect Reciprocity and the Evolution of Prejudicial Groups. *Scientific Reports, 8*(1), 13247. doi:10.1038/s41598-018-31363-z

Yoeli, E., Hoffman, M., Rand, D. G., & Nowak, M. A. (2013). Powering up with indirect reciprocity in a large-scale field experiment. *Proceedings of the National Academy of Sciences, 110*(Supplement 2), 10424. doi:10.1073/pnas.1301210110