A Unified Framework of Direct and Indirect Reciprocity

Supplementary Information:
A unified framework of direct and indirect reciprocity

Laura Schmid[1], Krishnendu Chatterjee[1], Christian Hilbe[2], Martin A. Nowak[3]
[1]IST Austria, Am Campus 1, 3400 Klosterneuburg, Austria
[2]Max Planck Research Group Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, 24306 Ploen, Germany
[3]Department of Organismic and Evolutionary Biology, Department of Mathematics, Harvard University, Cambridge MA 02138, USA

In this work, we introduce a model that combines direct and indirect reciprocity within a unified framework. Players do not only react to the games they are involved in. They also observe how their co-players act in interactions with third parties. We use this model to explore how individuals make use of different sources of information, and to compare how direct and indirect reciprocity facilitate the spread of cooperation throughout a population.

In the following, we describe the framework and the employed methods in detail. In **Section 1**, we begin by summarizing the previous literature on direct and indirect reciprocity. **Section 2** introduces our baseline model, for which we assume that players use simple reactive (first-order) strategies. In **Section 3**, we derive explicit formulas for the players' payoffs for any given population composition. **Section 4** offers an equilibrium analysis. We show under which conditions reactive strategies suffice to sustain a fully cooperative equilibrium. We prove that the respective conditions are stringent: if reactive strategies cannot sustain cooperation, more complex strategy classes cannot sustain it either. In **Section 5** we study the co-evolution of direct and indirect reciprocity among reactive players. We show that indirect reciprocity is most likely to evolve when there are only a few interactions, information is reliable, and mutations are not too abundant. In addition, we provide a framework for the evolution of strategies when individuals are able to use a mixture of direct and indirect reciprocity. **Section 6** discusses several model extensions. We explore how our model can capture different kinds of errors and incomplete information. We also discuss an alternative implementation of indirect reciprocity, according to which players can choose to ignore any direct information they may have. Finally, we incorporate higher-order strategies. The **Appendix** contains the proofs of our analytical results and the code for our evolutionary simulations.

1 Related literature

1.1 Previous literature on direct reciprocity.

There is by now an extensive literature on direct reciprocity, that is, the evolution of cooperation in repeated games. This literature has suggested various strategies that succeed in evolutionary simulations and tournaments[1–11], and it has discussed under which conditions cooperation can be evolutionarily stable[12–18]. Moreover, it has explored how the evolution of cooperation depends on model parameters, such as the players' memory[19–22] or which strategies players have access to[23–25]. A general summary of this literature can be found in recent reviews[26,27].

In the context of our paper, the literature most relevant is the recent work on zero-determinant (ZD) strategies for repeated games[28–42]. For the repeated prisoner's dilemma, Press and Dyson[28] have shown that a player can use such strategies to unilaterally enforce a linear relationship between the players' payoffs. A special case of these zero-determinant strategies are *equalizers*[43], with which a player can enforce that any co-player will get some fixed payoff, independent of the co-player's strategy. In our work, we use such equalizer strategies to construct Nash equilibria for both direct and indirect reciprocity.

Our work is also related to a previous study on crosstalk in repeated games[44]. Under crosstalk, players pay forward somebody's cooperation: if Alice helps Bob, this may increase the chance that Bob helps some third unrelated player Charlie, even though Bob has no prior positive experiences with that player. This form of generalized reciprocity is maladaptive: crosstalk undermines cooperation, and is disfavored to evolve in the first place. Herein, we adapt some of the mathematical techniques used in that paper[44] to derive a general payoff formula for direct and indirect reciprocity.

Our study adds to the previous literature on direct reciprocity in the following way:

(*i*) We fundamentally generalize the theory of ZD strategies, by extending it to settings in which players are unlikely to ever meet again, and in which direct reciprocity fails to maintain cooperation.

(*ii*) We explore in which scenarios players engage in direct reciprocity in the first place. To this end, we analyze for which parameters players learn to ignore any indirect information they might have about their present co-player.

1.2 Previous literature on indirect reciprocity.

Indirect reciprocity is an alternative mechanism for cooperation[45]. It can sustain cooperation even when players only interact once with each other, such that direct reciprocity is infeasible[46–48]. Much of the work has focused on the question how complex strategies need to be in order to establish stable cooperation[49–59]. This work has been summarized recently[60–62].

Since the influential work of Ohtsuki and Iwasa[53,54], it is a widely shared belief that successful strategies of indirect reciprocity need to be sufficiently complex. The complexity of strategies is typically evaluated in terms of how much information is needed to assess the reputation of a co-player. When players use *first-order strategies*, they assess a player purely based on the players' actions. For example, simple scoring assigns a good reputation to players who cooperated in their last interaction, and it assigns a bad reputation to players who defected[63]. *Second-order strategies* additionally depend on the reputation of the recipient. For example, according to the 'Stern' strategy[55], a player who defects against a good co-player deserves a bad reputation, whereas a player who defects against a bad co-player deserves a good reputation. In addition, *third-order strategies* take the actors' original reputation into account. For example, when players adopt the Staying strategy[64], a good player should keep his good reputation no matter how he treats a co-player who is deemed bad. By systematically exploring all deterministic third-order strategies, Ohtsuki and Iwasa have shown that there are only eight strategies that yield stable cooperation, called the 'leading eight'[53,54]. None of these eight strategies is first-order.

Most relevant to our study, Ohtsuki[65] has explored the adaptive dynamics of stochastic first-order strategies. He shows that the dynamics admits a fixed point in which everyone cooperates. However, since *ALLD* is the only locally stable fixed point, he concludes that no first-order strategy can sustain cooperation. Some subsequent studies have suggested that first-order strategies may suffice when they record more than a player's last action[66–68]. For example, in the recent paper by Clark *et al*[68] players count how often each other population member has defected so far. To sustain at least partial cooperation in the population, the paper suggests an innovative strategy called *GrimK*. Players with that strategy cooperate provided that both players' defection record is below K. For games in which there is a coordination-motive to cooperation, they show that one can always find a K such that *GrimK* is a strict Nash equilibrium.

Our study adds to this literature on indirect reciprocity in the following way:

(*i*) We prove that cooperation can be sustained in a Nash equilibrium when players use Generous Scoring. *GSCO* has an intuitive interpretation, it only depends on the co-player's very last action, and it is robust with respect to various kinds of errors (see **Section 6.1**).

(*ii*) Our results continue to hold when players differ in the information they have about each co-player, which has been a major obstacle for cooperation in some previous models of indirect reciprocity[69].

(*iii*) We show through simulations that the strategy dynamics in finite populations exhibits cycles. *ALLD* is typically invaded through conditionally cooperative strategies, which in turn are invaded by even more cooperative strategies.

1.3 Literature that combines elements of direct and indirect reciprocity.

There is only a handful of studies that explore how direct and indirect reciprocity interact. In an early study on the subject, Raub and Weesie explore the effectiveness of reciprocity when players are placed on a lattice[70]. They explore the stability of cooperation for three scenarios, differing in whether players do or do not receive information about their co-players' interactions with third parties. To this end, they analyze whether Grim/Trigger is an equilibrium, assuming there are no implementation or perception errors. In addition they explore the case that third-party information is received with some time lag. The study finds that immediate information about third-party interactions is most favorable to cooperation.

Pollock and Dugatkin propose a strategy for the repeated prisoner's dilemma called 'Observer Tit For Tat' (*OTFT*)[71]. Against a co-player with a joint previous history of play, *OTFT* behaves the same way as a *TFT* player. However, against an unknown co-player, *OTFT* takes into account third-party information. The paper explores the static competition between three strategies, *OTFT*, *TFT*, and *ALLD*. It is shown that *OTFT* can be stable against these three strategies even if the continuation probability approaches zero. However, for larger continuation probabilities, *TFT* is shown to be superior.

Roberts presents simulations for a meta-population setup when players can choose among a finite set of strategies[72]. The strategy set represents a selection from the direct and indirect reciprocity literature, and it includes first-order strategies (scoring) as well as second-order strategies (standing). When there are only few interactions between each pair of players, Roberts observes that most players adopt strategies of indirect reciprocity. This trend towards indirect reciprocity is even stronger when players have access to standing strategies. Importantly, however, the study assumes that there is public information about each player's reputation (i.e., it requires that all players agree on a given co-player's reputation at any point in time). For noisy and private information, most previously considered higher-order strategies of indirect reciprocity fail to maintain cooperation[69].

Finally, our work is related to two previous studies in which players are able to misrepresent their own reputation[73,74].

The first study is by Nakamaru and Kawata[73]. They consider a setup where players engage in two kinds of interactions. First, players interact in a series of prisoner's dilemma games. As in our study, players can decide whether to cooperate or defect, depending on the reputation of the opponent. Second, players are matched in pairs to communicate which reputation they assign to every population member. In particular, the model considers the case of private information – different individuals may assign different reputations to the same co-player. Compared to our study, strategies are more complex. Players do not only need to determine what they do in the prisoner's dilemma. They also need to specify to which extent they participate in rumour exchange, and whether or not they initiate wrong rumours about themselves. To this end, the study considers typical archetypes of strategies. For example, 'liars' defect in the prisoner's dilemma, and they misrepresent themselves as cooperative. In contrast, 'Advisors' are conditionally cooperative in the prisoner's dilemma. In the rumour exchanges, they spread true rumour about those co-players who defected against them. Finally, TFT-like players ignore rumours and just implement the traditional tit-for-tat strategy. The strategy dynamics is explored through computer simulations with up to 39 different strategies. Defectors win if players interact on average for one round or less. When there are slightly more interactions, rumour-based strategies like Advisor can succeed. Finally, with many pairwise interactions, TFT-like strategies persist.

The second study by Seki and Nakamaru[74] considers a similar setup as the first. However, this time strategies are represented differently. Each player's strategy is encoded as a list of numbers. One number represents how the player acts in the prisoner's dilemma. Two further numbers represent under which circumstances the player would spread positive or negative rumour, respectively. Another number represents under which condition the player would take third-party rumour into account into his own assessment of a player. This decision may depend on the reputation of the person who communicates the respective rumour. With four further numbers, the player represents to which extent direct information (positive or negative) or indirect information (positive or negative) affect the respective co-player's reputation. This reputation is measured by a number between -5 and 5. In addition to the liars whose aim is to self-advertise themselves, this study now also involves players who aim to destabilise the reputation system altogether. For example, such players may spread good rumours about everyone, which adds further noise to the system. The relative weight that players attribute to indirect information does not evolve. However, through extensive computer simulations, the paper shows that the more noise defectors introduce, the more difficult it becomes for cooperation to evolve through indirect reciprocity. Strategies that put more weight on direct reciprocity are more effective under these conditions.

Our study adds to this literature in the following way:

(*i*) We provide a general framework to explore the interaction of direct and indirect reciprocity. This framework allows us to systematically explore the evolution of reciprocity when players can choose among all possible first-order strategies, including stochastic strategies. It can easily be extended to include more complex strategies (see **Section 6.4**).

(*ii*) By focusing on comparably simple strategies in the main text, we allow for a more transparent comparison between direct and indirect reciprocity. In particular, we are able to derive analytical conditions when each kind of reciprocity is stable.

(*iii*) Simulation results necessarily depend on which strategies have competed. Adding further strategies can sometimes change the respective conclusions. In contrast, our analytical results are robust. When we find a cooperative strategy to be an equilibrium, it is stable against any possible deviation, even if more complex strategies become available.

2 Model description

We consider a well-mixed population of n players. Interactions take place in discrete time. In each time step, two players are chosen from the population at random to engage in one round of the prisoner's dilemma. The two chosen players independently decide whether to cooperate (C) or to defect (D). A cooperator pays a cost $c>0$ which yields a benefit $b>c$ to the co-player. Hence, the possible payoffs are $b-c$ if both players cooperate, $-c$ if only the focal player cooperates, b if only the co-player cooperates, and 0 if both defect. After each game, the process iterates with probability d. That is, with probability d, again two players are chosen at random to interact in a prisoner's dilemma. Otherwise, with probability $1-d$, the game terminates. Upon termination, the players' payoffs are calculated by averaging over all interactions in which they participated in.

To model how a player forms and updates her opinions about other group members, we consider players who use separate finite state automata to represent each other group member's reputation. In the baseline model, we assume each automaton only has two states, which we denote by G (the player considers the respective group member to be 'good') and B (the respective group member is considered 'bad'). Players cooperate with those group members they consider as good, and they defect against the bad ones. Such a binary representation of the co-player's current social standing has become standard in the literature on indirect reciprocity[53,60,61]. However, in **Section 6.3** we illustrate how our framework can be extended to allow for more nuanced representations of the co-player. With this extended framework, we can capture previous models in which players have a third 'neutral'[75] or 'unknown'[76] state, or models in which the co-player's score is measured in integer values[49,50]. The extended framework also allows us to capture situations in which a player's assessment of a co-player depends on more than the co-player's last observed action[19,66].

While a player can be in different states with respect to different group members, each player uses a uniform rule to update all his automata. We interpret this rule as the player's strategy. In the baseline model, players use simple reactive strategies. The strategy of player i is represented by a vector $(y_i, p_i, q_i, \lambda_i) \in [0,1]^4$. The first entry y_i is the probability that player i initially assigns a good reputation to a given co-player (without having observed any interaction of that co-player before, see **Fig. 1d**). We assume that all of the player's $n-1$ automata are initialized independently. In particular, for $0<y_i<1$, the player may assign different initial states to different co-players. However, all our qualitative results remain unchanged if a player's initial assignments are fully correlated, such that the player assigns the same initial reputation to all other group members. The values of p_i and q_i determine how the player updates a co-player's state after a direct interaction (**Fig. 1e,f**). A cooperative co-player is assigned a good reputation with probability p_i, whereas a defecting co-player is assigned a good reputation with probability q_i. As an example, the strategy *ALLD* sets $y=p=q=0$, whereas *ALLC* uses $y=p=q=1$.

Finally, the last parameter λ encodes to which extent players take indirect information into account when assigning reputations to others (**Fig. 1g**). To model the impact of indirect information, we assume that all players can observe the interactions of all other population members (for the case of incomplete information, see **Section 6.1**). After player i witnesses an interaction between player j and some third party, i updates player j's reputation with probability λ_i. If player j's reputation is updated, the new reputation is good with probability p_i if player j has cooperated with the third party, and it is good with probability q_i if player j has defected. We refer to λ as the player's *receptivity*, as it controls to which extent the player is receptive to indirect information.

In the special case that all players set $\lambda=0$, they completely ignore third party interactions. We refer to this case as 'direct reciprocity'. In the other limit $\lambda=1$, a player equally takes into account all actions of the other group members, no matter whether the player is directly involved. We refer to this case as 'indirect reciprocity'. We note that even a player with $\lambda = 1$ may occasionally cooperate based on her direct experience with the given co-player. Such an instance occurs for example if the same two players are randomly chosen to interact twice in a row. In that case, a player's behavior in the second round will depend on what happened in the first. This seems natural: in most applications, even a player who routinely takes into account third-party information would not ignore any piece of information merely because it stems from a direct encounter. Nevertheless, it can be useful from a conceptual perspective to consider an alternative model of indirect reciprocity, where all decisions are solely based on third-party information. We consider such a model in **Section 6.2**.

The archetypal strategy of direct reciprocity, Tit-for-Tat, corresponds to the vector *TFT*$=(1,1,0,0)$. The analogous strategy in the indirect reciprocity literature, simple scoring[63], is given by *SCO*$=(1,1,0,1)$. In **Extended Data Fig. 1a–d**, we provide a graphical illustration of our model and the resulting reputation dynamics. **Table S1** gives a summary of the parameters of our model.

Previous work has shown that the evolution of reciprocity is sensitive to the presence of noise[69,77,78]. We thus assume that observations may be subject to perception errors. In this way, we allow indirect information to be less reliable than direct information. Specifically, we assume that when observing an indirect interaction, player i misinterprets somebody's cooperation as defection with probability $\varepsilon<1/2$. Similarly, a player's defection may be taken as cooperation with the same probability. For simplicity direct interactions are not subject to perception errors in the baseline model. However, in **Section 6.1**, we analyze a model extension that includes perception and implementation errors for both modes of reciprocity.

We note that the strategies in this baseline model only make use of first-order information[52,54]. A

	Parameter	Interpretation
Fixed Parameters	n	Population size
	b, c	Benefit and cost of cooperation
	ε	Error rate for indirect information
	d	Probability of another interaction in the entire population
	δ	Probability of another interaction between a given pair of players, introduced in **Section 4**
	μ	Mutation rate, introduced in **Section 5**
	β	Selection strength, introduced in **Section 5**
Evolving traits	y	Probability to assign a good reputation to unknown players
	p	Probability to assign a good reputation to co-players who cooperate
	q	Probability to assign a good reputation to co-players who defect
	λ	Probability to use indirect information

Table S1: Parameters of the model. Our model involves a number of fixed parameters that are the same for all players and kept constant over time. In addition, our model considers four evolving traits. The values of the evolving traits may differ between individuals. They are kept constant over the course of a game, but they may change over an evolutionary timescale (see **Section 5**).

player's reputation only depends on the player's action. It does not depend on the standing of the recipient, or on the player's previous reputation. The assumption of first-order strategies greatly facilitates the calculation of the players' payoffs in **Section 3**. However, in **Section 6.4** we revisit higher-order strategies, and we discuss how they can be captured by our framework.

Finally, we note that starting with the influential work of Ohtsuki and Iwasa[53,54] much of the literature on indirect reciprocity considers the case of public information[55,58,64,79]. In such models, players do not observe each others' interactions with third parties directly. Rather there is a central observer who monitors all interactions in a population, assigns new reputations, and disseminates the updated reputations to all population members. Models of public information have the useful mathematical property that all players agree on the reputation of any other population member (because everyone receives the same information from the same source). In contrast, herein we are interested in the formation of reputations when some players base their decisions on direct interactions, whereas others may also take indirect information into account. In such a situation, players may no longer agree on the reputation they assign to some given co-player. Our model is thus – necessarily – a model of private information[69,73,74,78].

3 A unified payoff equation for direct and indirect reciprocity

In this section we derive an explicit expression of the payoffs when all players use reactive strategies (y, p, q, λ). To this end, let $x_{ij}(t)$, be the probability that player i assigns a good reputation to co-player j at time t. As shown in the **Methods** section of the main text, this quantity satisfies the recursion,

$$\begin{aligned} x_{ij}(t{+}1) = {} & (1{-}\bar{w})\, x_{ij}(t) \\ & + w\left(x_{ji}(t)\, p_i + (1{-}x_{ji}(t))\, q_i\right) \\ & + (\bar{w}{-}w)\,(1{-}\lambda_i)\, x_{ij}(t) \\ & + w\lambda_i \sum_{l\neq i,j} \Big(\big(1{-}\varepsilon\big) x_{jl}(t) {+} \varepsilon\big(1{-}x_{jl}(t)\big)\Big)\, p_i + \Big(\big(1{-}\varepsilon\big)\big(1{-}x_{jl}(t)\big) + \varepsilon x_{jl}(t)\Big)\, q_i, \end{aligned} \tag{1}$$

with

$$x_{ij}(0) = y_i. \tag{2}$$

The parameter $\bar{w} = 2/n$ is the probability that a particular player is chosen to interact in the next round, and $w = 2/\big(n(n{-}1)\big)$ is the probability that a particular pair of players is chosen. To obtain an expression for the payoffs, we take Eq. (1), collect all terms with $x_{kl}(t)$, and define $r_i := p_i - q_i$, which yields

$$\begin{aligned} x_{ij}(t{+}1) = {} & \Big(1{-}w{-}\lambda_i(\bar{w}{-}w)\Big) x_{ij}(t) \ + \ w\, r_i\, x_{ji}(t) \ + \ w\lambda_i(1{-}2\varepsilon) r_i \sum_{l\neq i} x_{jl}(t) \\ & + \Big(w\, q_i + \lambda_i(\bar{w}{-}w)(\varepsilon r_i + q_i)\Big). \end{aligned} \tag{3}$$

Eq. (3) indicates that the value of $x_{ij}(t{+}1)$ is a linear function of the respective probabilities $x_{kl}(t)$ in the previous round.

For further manipulation, it is useful to rewrite Eq. (3) using matrix notation. To this end, we collect the players' probabilities to assign a good reputation to their co-players in an $n(n-1)$-dimensional column vector,

$$\mathbf{x}(t) := \big(x_{12}(t), \ldots, x_{1n}(t);\ x_{21}(t), \ldots, x_{2n}(t);\ \ldots\ ;\ x_{n1}(t), \ldots x_{n(n-1)}(t)\big)^{\mathsf{T}}. \tag{4}$$

Similarly, we collect the factors in the first line of Eq. (3) in an $n(n{-}1) \times n(n{-}1)$ matrix $\mathbf{M} = (m_{ij,kl})$, with entries

$$m_{ij,kl} = \begin{cases} 1{-}w{-}\lambda_i(\bar{w}{-}w) & \text{if } i = k \ \text{ and } j = l \\ w\, r_i & \text{if } i = l \ \text{ and } j = k \\ w\lambda_i(1{-}2\varepsilon) r_i & \text{if } i \neq l,\ j \neq l, \ \text{ and } j = k \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Finally, we collect the constant term in the second line of Eq. (3) in an $n(n-1)$-dimensional column vector $\mathbf{v} = (v_{ij})$ with entries

$$v_{ij} = w\, q_i + \lambda_i(\bar{w} - w)(\varepsilon r_i + q_i) \quad \text{for all } j. \tag{6}$$

Using this notation, we can write Eq. (3) as $\mathbf{x}(t+1) = \mathbf{M}\mathbf{x}(t) + \mathbf{v}$. Based on this equation, we calculate the weighted average

$$\mathbf{x} := (1-d)\sum_{t=0}^{\infty} d^t \cdot \mathbf{x}(t) = (\mathbf{1} - d \cdot \mathbf{M})^{-1}\big((1-d)\mathbf{x_0} + d\mathbf{v}\big). \tag{7}$$

Here, $\mathbf{1}$ denotes the identity matrix, and $\mathbf{x_0}$ is the shorthand notation for $\mathbf{x}(0)$ with entries as defined by Eq. (2). The $n(n-1)$ entries x_{ij} of this vector $\mathbf{x}$ can be interpreted as the probability to find player i's automaton with respect to j in the good state in a randomly picked round. We use Eq. (7) to compute the expected payoff π_i of player i as

$$\pi_i = \frac{1}{n-1}\sum_{j \neq i}(x_{ji}\, b - x_{ij}\, c). \tag{8}$$

In **Extended Data Fig. 10**, we compare this analytically derived payoff with the payoff obtained from simulations of the game dynamics. For these simulations, we consider a population where $n-1$ players use the conditionally cooperative strategy $\sigma_C = (1, 1, q, \lambda)$, whereas the remaining player is a defector, $\sigma_D = (0, 0, 0, \lambda)$. We compute and simulate the payoffs for different values of λ for two different scenarios (with different continuation probabilities and error rates). In all cases, we observe that the analytically derived payoffs fully match the simulation results.

When using the above equations to calculate payoffs, the computationally most expensive step is to find the inverse of the $n(n-1) \times n(n-1)$ matrix $(\mathbf{1} - d\mathbf{M})$ in Eq. (7). This computation can be made more efficient when several players in a population adopt the same strategy (which will often be the case in evolutionary simulations). In that case, one can exploit the symmetries of a well-mixed population: all players who adopt the same strategy are expected to receive the same payoff. For the computation of payoffs it is thus not necessary to distinguish between all n players. It suffices to distinguish between all strategies that are present in the population – a usually far smaller number. In the **Appendix**, we show how payoffs can be computed more efficiently when these symmetries are taken into account.

4 Equilibrium analysis

4.1 Characterization of all reactive Nash equilibria

In the following, we wish to characterize which reactive strategies $\sigma_i = (y_i, p_i, q_i, \lambda_i)$ are Nash equilibria (in the next subsection we will then focus on those Nash equilibria that yield full cooperation). A strategy is a Nash equilibrium if no player has an incentive to unilaterally deviate from it. To simplify the notation, and to make our results comparable to the previous literature on direct and indirect reciprocity, it is useful to introduce an additional parameter δ, which is the continuation probability for a pair of players.

Lemma 1. *Consider a population of size n, and let d be the probability that another random pair is drawn from the population after the current round. Let δ be the probability that a given pair of players interacts again after it has just participated in an interaction. Then*

$$\delta = \frac{2d}{2d + (n-1)n(1-d)}. \tag{9}$$

All proofs are provided in the **Appendix**. While d is the global continuation probability (the probability that another game will be played in the entire population), the parameter δ is the continuation probability for each pair of players. We note that the above formula implies that $\delta = 1$ if and only if $d = 1$, and that $\delta = 0$ if and only if $d = 0$, as one may expect. Using the pairwise continuation probability δ, we can formulate the following result that will help us to characterize the set of all Nash equilibria.

Lemma 2. *Consider a homogeneous population with strategy $\sigma = (y, p, q, \lambda)$, and let $r := p - q$. Then on average, players assign a good reputation to each other with probability*

$$x = \frac{\Big(1-\delta\Big)y + \delta\Big(q + \lambda(n-2)(q+r\varepsilon)\Big)}{\Big(1-\delta\Big) + \delta\Big(1-r + \lambda(n-2)(1-r+2r\varepsilon)\Big)} \tag{10}$$

In particular, for a generic game with $n > 2$, $\varepsilon > 0$, and $0 < \delta < 1$ we have:

1. *The population is fully cooperative (all players' automata are in the G state for the entire game) if and only if $y = p = 1$ and either $\lambda = 0$ or $q = 1$.*
2. *The population is fully defecting (all players' automata are in the B state for the entire game) if and only if $y = q = 0$ and either $\lambda = 0$ or $p = 0$.*

Finally, the following lemma describes which payoff a single player can get from deviating from the *resident* strategy (the strategy everyone else in the population applies). For brevity, we will sometimes refer to the deviating player as the *mutant*.

Lemma 3. *Consider a population where all but one player apply the resident strategy $\sigma = (y, p, q, \lambda)$. Let $r := p - q$. Then the mutant's payoff π' takes the form*

$$\pi' = A_1 + A_2(r - r^*_\lambda)x'. \tag{11}$$

Here, x' is the mutant's average cooperation rate against the residents, $A_1, A_2 > 0$ are constants that only depend on the resident strategy, and

$$r^*_\lambda = \frac{1 + (n-2)\delta\lambda}{1 + (n-2)(1-2\varepsilon)\lambda} \cdot \frac{c}{\delta\, b}. \tag{12}$$

According to Lemma 3, the payoff of the mutant is a linear function of the mutant's cooperation rate. Due to the properties of linear functions, it follows that the mutant's payoff is either maximized by choosing $x'=1$ (if $r \geq r^*_\lambda$), or by choosing $x'=0$ (if $r \leq r^*_\lambda$). That is, for a resident strategy σ with $r > r^*_\lambda$, *ALLC* is a best response. In analogy to the case of reactive strategies for the repeated prisoner's dilemma[80], we call the set of all such strategies $\sigma \in [0,1]^4$ the *cooperation rewarding zone*. Conversely, for any strategy σ with $r < r^*_\lambda$, *ALLD* is a best response, yielding the *defection rewarding zone*. In between these two zones, for $r = r^*_\lambda$, any mutant strategy obtains the same payoff $\pi' = A_1$. Strategies for which $r = r^*_\lambda$ are called *equalizers*. Equalizer strategies have been previously described in models of direct reciprocity[28,43,81]. Lemma 3 guarantees that analogous strategies also exist when players take arbitrary amounts of indirect information into account.

It is important to note that Lemma 3 makes no restrictions on the mutant strategy. For the lemma to hold, we do not require the mutant to choose a strategy of the form $\sigma' = (y', p', q', \lambda')$. Instead the mutant may take arbitrarily many past actions of the co-player into account, and she may combine direct and indirect information in non-trivial ways. According to Lemma 3, the mutant's eventual payoff is solely determined by her resultant average cooperation rate.

Based on the above lemmas, we can now characterize all Nash equilibria among the reactive strategies.

Theorem 1 (Characterization of Nash equilibria)**.**
Consider $0 < \delta < 1$*, and a strategy* $\sigma = (y, p, q, \lambda)$.

1. In a game with $n > 2$ *and* $\varepsilon > 0$*, a strategy* σ *with* $\lambda > 0$ *is a Nash equilibrium if and only if it is either ALLD or an equalizer strategy,*

$$y = p = q = 0, \quad \text{or} \quad p - q = r^*_\lambda. \tag{13}$$

We refer to strategies of the form (13) as generic Nash equilibria.

2. If $\lambda = 0$*,* $n = 2$ *or* $\varepsilon = 0$*, the strategy* σ *is a Nash equilibrium if and only if it is either generic, or if one of the following two cases applies,*

$$y = q = 0, \;\; p < r^*_\lambda \quad \text{or} \quad y = p = 1, \;\; q < 1 - r^*_\lambda. \tag{14}$$

Remark 1. We emphasize that while Theorem 1 characterizes the Nash equilibria among the reactive strategies, it does not restrict the strategies deviating players may employ. The Nash equilibria described in Theorem 1 are robust against *any* possible mutant strategy, including mutant strategies that take higher order information into account, or mutant strategies that depend on the whole history of previous play.

4.2 Cooperative Nash equilibria

In the following, we are interested in those strategies that can sustain high levels of cooperation in a population. To this end, we call a strategy σ a *cooperative Nash equilibrium* if (*i*) it constitutes a generic Nash equilibrium, and

(*ii*) the cooperation rate in a homogeneous σ-population approaches one as the error rate ε goes to zero.

From Eqs. (10) and (13), it follows that cooperative Nash equilibria need to be equalizers of the form $\sigma=(1,1,q^*_\lambda,\lambda)$ with $q^*_\lambda := 1-r^*_\lambda$ and r^*_λ defined by Eq. (12). In the case of direct reciprocity ($\lambda=0$), the corresponding cooperative Nash equilibrium thus takes the form

$$y=1, \quad p=1, \quad q^*_0 = 1-\frac{c}{\delta\, b}. \tag{15}$$

This strategy has been described earlier and is known as *Generous Tit-for-Tat* (*GTFT*)[2,3]. Surprisingly the analogous strategy for indirect reciprocity ($\lambda=1$) has not been described before, given by

$$y=1, \quad p=1, \quad q^*_1 = 1-\frac{1+(n-2)\delta}{1+(n-2)(1-2\varepsilon)}\,\frac{c}{\delta\, b}. \tag{16}$$

In analogy to the previous case, we call this strategy *Generous Scoring* (*GSCO*). These two Nash equilibria do not need to exist for all parameter values because the respective value of q^* may become negative. Theorem 2 summarizes the necessary and sufficient conditions for cooperative Nash equilibria to exist.

Theorem 2 (Existence of cooperative Nash equilibria)**.**

1. There is a cooperative Nash equilibrium $\sigma = (1,1,q^*_0,0)$ *in which players exclusively use direct information if and only if* $\delta \ge \delta_0$ *with*

$$\delta_0 = \frac{c}{b}\,. \tag{17}$$

2. There is a cooperative Nash equilibrium $\sigma=(1,1,q^*_1,1)$ *in which players use indirect information if and only if* $2\varepsilon < 1-c/b+1/(n-2)$*, and* $\delta\ge\delta_1$ *with*

$$\delta_1 = \frac{c}{b+(n-2)\big((1-2\varepsilon)b-c\big)}. \tag{18}$$

3. For $0<\lambda<1$*, there is a cooperative Nash equilibrium* $\sigma=(1,1,q^*_\lambda,\lambda)$ *if and only if there is a cooperative Nash equilibrium for* $\lambda=0$ *or* $\lambda=1$*.*

Theorem 2 gives three major insights:

First, for sustaining cooperation in a Nash equilibrium there is no advantage of using both direct and indirect information simultaneously (i.e., to choose $0<\lambda<1$). From the third part of Theorem 2 it follows that if some cooperative Nash equilibrium in reactive strategies exists at all, there is always a cooperative Nash equilibrium for either $\lambda=0$ or $\lambda=1$.

Second, the first two parts of Theorem 2 suggest that for $\delta_0, \delta_1 \in [0,1]$ we can partition the parameter space of the game into four distinct regions:

(*i*) When $\delta < \min\{\delta_0, \delta_1\}$, there is no cooperative Nash equilibrium;

(*ii*) When $\delta_0 < \delta < \delta_1$, full cooperation can be sustained with direct but not with indirect reciprocity;

(*iii*) When $\delta_1 < \delta < \delta_0$, cooperation can be sustained with indirect but not with direct reciprocity;

(*iv*) When $\delta > \max\{\delta_0, \delta_1\}$ both direct and indirect reciprocity allow for stable cooperation.

We provide a graphical illustration of these parameter regions as **Fig. 2c** in the main text.

Third, parameter changes affect the feasibility of cooperation as follows.

Corollary 1. *If a cooperative Nash equilibrium exists for given values n, δ, and ε, then there also exists a cooperative Nash equilibrium for any $n' \geq n$, $\delta' \geq \delta$, and $\varepsilon' \leq \varepsilon$.*

In a nutshell, the above corollary states that the stability of cooperation is monotonic in the population size, the pairwise continuation probability, and the error rate. All other parameters kept equal, the conditions for a cooperative Nash equilibrium are easier to meet if the population size is larger, if there are more rounds to be played in expectation, or if there is less noise on indirect information.

Several remarks are in order.

1. While the above results for $\lambda = 0$ fully recover previous results on direct reciprocity among reactive players[2,3], the results on indirect reciprocity ($\lambda = 1$) may come as a surprise. Previous work on indirect reciprocity suggests that reactive ('first-order') strategies are unable to sustain cooperation[50,51,53]. The intuition for this pessimistic result is that in a resident population of perfect discriminators (with $y = p = 1$, $q = 0$), a resident may have no incentive to discriminate against a rare *ALLD* mutant. By punishing a defector, discriminators harm their own reputation, which puts them at risk to receive less cooperation in future. Our model circumvents this problem by allowing for stochastic strategies. By increasing their q-value from zero, discriminators are able to reduce the effective cost of punishment. Once they reach $q = q^*_\lambda$, the expected long-term loss in reduced reputation after defecting against an *ALLD* mutant exactly matches the short-term gain in saved cooperation costs. In the Nash equilibrium, discriminators no longer bear an effective cost of punishing defectors.

2. The strategies Generous Tit-for-Tat and Generous Scoring are Nash equilibria, but they are not evolutionarily stable. That is, if the entire population adopts one of these strategies, no mutant is favored to invade, but mutants are not opposed to invade either. Evolutionary stability is generally difficult to achieve in repeated interactions. For the repeated prisoner's dilemma, it has been shown that no strategy is evolutionary stable in the absence of errors[12,13]. Any resident strategy can be invaded by 'stepping-stone' mutations, which in turn may facilitate the invasion of further mutant strategies[18]. Simulations suggest that cooperation comes and goes in cycles; the frequencies of cooperative and defecting strategies oscillate over time[26]. However, how often cooperative strategies are played over an evolutionary timescale critically depends on whether or not cooperative Nash equilibria exist[17]. We will revisit this issue in the next section.

For indirect reciprocity it has been suggested that evolutionary stability is feasible if more complex strategies are permitted[53,54]. The respective 'leading-eight' strategies require that the reputation of a player does not only depend on a player's action, but additionally on the reputation of the opponent (and sometimes also on the player's own previous reputation). While these results of Ohtsuki and Iwasa have been tremendously important for understanding which properties stable norms ought to have, their framework is based on some rather restrictive assumptions. Most importantly, information transmission is assumed to be public. In particular, all population members agree on the reputation they assign to a given co-player. In contrast, we allow some players to use a co-player's public record to assign reputations, whereas others may make their judgments solely based on their direct experiences. As a consequence, different players may assign a different reputation to the same co-player. For such a private information scenario, it has been recently shown that not even the leading-eight strategies are able to sustain high cooperation rates in the presence of perception errors[69].

3. Previous results on the stability of cooperation with indirect reciprocity only show stability *within* the given strategy set considered[53,57,64,67]. This leaves it open whether more complex strategies could invade, either due to neutral drift or because of a selective advantage. In contrast, the equilibrium results presented in this section do not restrict which strategies mutants are permitted to take. The Nash equilibria that we describe are equilibria with respect to *all possible* mutant strategies (see Remark 1).

4. In the limit of rare errors, the conditions that we obtain for the feasibility of cooperation with direct and indirect reciprocity are strict, as the following result shows.

Theorem 3 (Optimality of equilibrium conditions)**.**
Consider a population of size n interacting in a population game as introduced in Section 2, and suppose $\varepsilon \to 0$. Then there exists a Nash equilibrium (not necessarily in reactive strategies) that yields full cooperation if and only if there is a cooperative Nash equilibrium in reactive strategies.

That is, our focus on simple reactive strategies does not restrict the player's ability to sustain cooperation. Theorem 3 says that if full cooperation is feasible at all in the limit of rare errors, it is already feasible with reactive strategies.

With the above equilibrium analysis we explore which strategies can maintain cooperation if the respective strategy is already adopted by a large majority of the population. Importantly, however, *ALLD* is also an equilibrium for all considered parameter values (Theorem 1). Therefore, even if cooperative equilibria exist, our results do not imply that evolving populations would settle at these equilibria. Before we discuss the corresponding evolutionary questions in the next section, we highlight a few reasons why we deem equilibrium analyses, like the one provided above, valuable: (*i*) The existence of cooperative equilibria can be considered a minimum requirement for cooperation to evolve. (*ii*) Our analytical equilibrium conditions in Theorem 2 allow predictions on how certain parameter changes affect the feasibility of cooperation (see Corollary 1). (*iii*) Even if we are to find that cooperative Nash equilibria have problems to evolve, this may have important policy implications. For example, such a finding could suggest that temporarily, additional incentives for cooperation are necessary.

5 Evolutionary analysis

5.1 Description of the evolutionary process

In the previous section, we have considered a static setup. We have explored which reactive strategies can sustain an equilibrium with full cooperation. In the following, we no longer assume that players settle at an equilibrium automatically. Instead players may choose arbitrary reactive strategies. Over time, they adapt their behaviors through social learning.

To model this adaptation dynamics, we consider a pairwise comparison process[82] in a well-mixed population of constant size n. Initially, all players are assumed to use some arbitrary strategy $(y_0, p_0, q_0, \lambda_0)$. Then in every time step of the evolutionary process, two possible events can happen.

Exploration event. With probability μ (the mutation rate), one of the players is chosen at random. All players have the same probability to be chosen. This player then adopts a randomly chosen new strategy (y', p', q', λ'). The new values for y', p', q' are drawn from a uniform distribution on the unit interval. For λ' we first focus on the case that players either use direct or indirect reciprocity, such that $\lambda' \in \{0, 1\}$. We discuss the case of intermediate values of λ' in **Section 5.4**.

Imitation event. With probability $1-\mu$, two players are randomly drawn from the population, a learner and a role model. Given the composition of the population, their expected payoffs π_L and π_R are calculated using Eq. (8). We assume the learner adopts the role model's strategy with probability

$$\xi = \frac{1}{1 + e^{-\beta(\pi_R - \pi_L)}} \tag{19}$$

The parameter $\beta \geq 0$ reflects the strength of selection. It determines to which extent the learner's decision depends on the role model's relative success. In the special case that $\beta = 0$, the imitation probability simplifies to $\xi = 1/2$, such that imitation occurs purely at random. As β becomes larger, the learner is increasingly biased to only imitate role models with a higher payoff.

The above two elementary events, exploration and imitation, give rise to an ergodic process on the space of all population compositions. We explore this process with computer simulations. For each simulation, we use a fixed set of parameter values. Simulations are run for sufficiently long, such that the population's average cooperation rate has converged and is independent of the initial population. In addition to this cooperation rate, we record which strategies the players use over time.

Such simulations are computationally expensive. Even when using optimized algorithms (see Section 7.1), they need to keep track of all s strategies in the population and to invert $s^2 \times s^2$ matrices at each selection event. To make these computations more efficient, we have employed the method of Imhof and Nowak[83] for some of our results. This method assumes that mutations are rare, $\mu \to 0$. In that case, once a mutant enters the population, this mutant can be expected to reach fixation or to go extinct before the next mutation arises[84,85]. As a result, at any given time at most two different strategies are present in the population. The mutant's fixation probability in a resident population with strategy σ_R can be calculated explicitly, using the formula[86]

$$\rho(\sigma_M, \sigma_R) = \frac{1}{1 + \sum_{i=1}^{n-1} \prod_{k=1}^{i} e^{-\beta\,(\pi_M(k) - \pi_R(k))}}. \tag{20}$$

Here, $\pi_M(k)$ and $\pi_R(k)$ are the respective payoffs when k individuals in the population employ the mutant strategy. Once the mutant has gone extinct, or taken over the population, the next mutation occurs. This rare-mutation limit allows for more efficient computation of the resulting dynamics. The saved computation time can then be used, for example, to run the process for a larger number of generations. In this way we can ensure that sufficiently many mutant strategies are drawn over the course of a simulation to obtain a representative sample of the entire strategy space.

Overall, the process generates a sequence of strategies $(\sigma_0, \sigma_1, \ldots)$ that the residents use after each introduced mutant strategy. Based on this sequence, we can again calculate the average cooperation rate and the average strategy traits that evolve for different parameter values.

5.2 Impact of parameters on the evolutionary results

As summarized in **Table S1**, our model includes a number of different parameters. Throughout most of the main text, we have focussed on two of them, (*i*) the game's continuation probability δ, and (*ii*) the probability ε that indirect observations are subject to perception errors. In addition, our model depends on the benefit-to-cost ratio b/c of cooperation; the population size n; the strength of selection β; and the mutation rate μ. In the following, we describe the effect of each of these parameters on the co-evolution of cooperation and reciprocity. In particular, we focus on two aspects. (1) How do changes in the respective parameter affect the evolution of cooperation? (2) How do changes in the respective parameter affect the proportion of players who use indirect reciprocity? In each case, we first use our static equilibrium results and the previous literature to form expectations on the impact of the respective parameter. After that, we compare these expectations with the actual behavior observed in simulations.

(i) Continuation probability δ.

Because the expected number of rounds between two given players is given by $1/(1-\delta)$, the parameter δ determines how often two players interact on average. Based on our static equilibrium analysis (Corollary 1 in Section 4), larger values of δ should facilitate the evolution of cooperation. With respect to the abundance of indirect reciprocity, our equilibrium analysis suggests that indirect reciprocity should be particularly favoured for intermediate values of δ, where direct reciprocity has difficulties to evolve. For sufficiently large continuation probabilities, our static equilibrium analysis does not allow us to make a clear prediction, because both *GTFT* and *GSCO* are equilibria for $\delta \to 1$.

We have performed two sets of simulations in which we have varied the continuation probability systematically, one for the limit of rare mutations (**Fig. 5d**) and one for a positive mutation rate (**Fig. 5h**). In both cases, our simulation results are in line with the above predictions. For any error rate ε, cooperation becomes more likely as we increase δ. Moreover, indirect reciprocity is most abundant when δ is intermediate. For $\delta \to 1$, we find that direct reciprocity is more likely to emerge. We provide some analytical insight into this last effect in Section 5.3.

(ii) Error rate ε on indirect observations.

Based on our equilibrium analysis, we would expect that an increase in ε tends to reduce cooperation (Corollary 1 in Section 4). This reduction may be less severe for large continuation probabilities, where players may switch to direct reciprocity instead. With respect to the abundance of indirect reciprocity, we expect errors to either have no effect (e.g., when no cooperation evolves and information is irrelevant), or to have a negative effect.

Again, there are two sets of evolutionary simulations in which we vary the error rate ε systematically, see **Fig. 5d,h**. In general, these simulations are in agreement with the above expectations. In case the error rate has an effect on cooperation at all (i.e., for intermediate δ), this effect is negative. Similarly, in those cases in which the error rate affects the player's propensity to use indirect reciprocity (i.e., for intermediate to large values of δ), this effect is again negative.

The next four parameters are not specific to models of reciprocity. They feature an important role in any evolutionary model of cooperation. Some of these parameters are known to have non-trivial effects on simulation outcomes. For example, it has been shown that strategy abundances may change non-monotonically in population size and selection strength[87–89]. Similarly, an increase in the system's mutation rate can change the evolutionary dynamics altogether[90]. Analytical predictions for these effects are usually only possible in simple systems with a few strategies. In the following, we therefore only provide some basic intuition.

(iii) Benefit-to-cost ratio b/c.

According to Theorem 2, a larger benefit-to-cost ratio is favorable to both, a *GTFT* equilibrium and a *GSCO* equilibrium. We would thus expect that an increase in b/c enhances cooperation. With respect to the abundance of indirect reciprocity, our equilibrium analysis does not make any predictions.

We explore the effect of the benefit-to-cost ratio on evolution in **Extended Data Fig. 3a,b**. There, we distinguish between three scenarios, depending on how many interactions occur on average and on how noisy third-party information is. In all scenarios, we find that cooperation is an increasing function in b/c. The effect of b/c on the abundance of indirect reciprocity is more irregular. However, for all benefit-to-cost ratios, we observe that an environment with intermediately many interactions and reliable information is more favorable to indirect reciprocity than an environment with many interactions and unreliable information.

(iv) Population size n.

The effect of n is more difficult to predict because population size affects the dynamics on multiple levels. First, according to Eq. (18), it affects the stability of conditional cooperation under indirect reciprocity. The intuition here is that the larger the population, the more likely players already have some third-party information when they encounter a given co-player for the first time. Increases in population size would therefore increase cooperation, because players can more easily avoid helping an unconditional defector. However, this effect should become weaker once the population is already sufficiently large. Second, the size of a population affects the evolutionary process. On the one hand, small populations tend to select for spiteful behaviors[87,91]. The intuition here is that the smaller a population, the easier it becomes for players to show an above-average performance by simply diminishing the payoff of all their co-players. On the other hand, the dynamics in small populations is more stochastic. Even if a strategy is slightly disfavored, it still has a reasonable chance to reach fixation due to stochastic effects if the population is sufficiently small.

Based on these mechanisms, we can expect the following behavior. For the smallest meaningful population size, $n = 2$, cooperation is disfavored by the evolutionary process (i.e., the cooperation rate is smaller than 1/2), because of the effects of spite. Moreover, since direct and indirect reciprocity are indistinguishable for $n = 2$, both should have equal frequency. While these predictions for $n = 2$ hold for all scenarios, the predictions for $n > 2$ will depend on the other model parameters. However, here again one may predict that scenarios with sufficiently many interactions tend to favor cooperation. Scenarios with intermediately many interactions and reliable information tend to favor indirect reciprocity.

Our simulations for varying population sizes indeed follow this general pattern (**Extended Data Fig. 3c,d**). We note that in comparably large populations ($n > 500$), cooperation rates are non-monotonic when simulations are run for a short time-span (green and orange curves in **Extended Data Fig. 3d**). In large populations, it takes on average more time until any given population is invaded. Therefore, the outcome of short simulations is more dependent on the initial population. Because the defection rewarding zone of our strategy space is larger than the cooperation rewarding zone (Section 4.1), non-cooperative populations are initially favored. This may explain the decline of cooperation for larger population sizes when simulations are run for a short time span.

(v) Selection strength β.

Selection strength only affects the evolutionary process. Therefore we cannot directly infer its effects based on our static equilibrium predictions. Still we can form some intuition in the following way. In the most simple case of $\beta = 0$, payoffs are irrelevant for the spread of a strategy. For such a neutral process, we expect the average cooperation rate and the average abundance of indirect reciprocity to approach 50%. For small but positive selection strengths, we would expect cooperation to be slightly disfavored in all scenarios. The intuition for this is as follows: Because selection is weak, all strategies are played with almost equal abundance. Because the best response to a majority of strategies is *ALLD* (Section 4.1), this effect is expected to give a slight advantage to non-cooperative strategies. For intermediate to strong selection, we expect the outcome to depend again on the scenario. In scenarios with sufficiently many interactions, cooperation can evolve. In addition, provided that cooperation evolves, we expect more indirect reciprocity when there are intermediately many interactions and information is reliable.

We systematically vary selection strength in the simulations shown in **Extended Data Fig. 3e,f**. Again, the results seem to generally agree with the intuition provided above. As with population size, we find that the effect of selection strength on cooperation can be non-monotonic. However, this time, we observe this non-monotonicity even in the case that simulations are run for a long time (orange curve in **Extended Data Fig. 3e**; for the respective simulations, we have consecutively introduced $5\cdot 10^7$ mutant strategies into the population. We have checked with additional simulations that 10^8 mutant strategies yield the same output).

In **Extended Data Fig. 4**, we explore this non-monotonicity in more detail. Contrary to what one may expect, we do not observe that this non-monotonicity is due to a reduced robustness of cooperators (**Extended Data Fig. 4a**). Moreover, mutants that succeed in invading a conditionally cooperative resident are highly cooperative themselves (**Extended Data Fig. 4d**). As a result, highly cooperative populations actually become more abundant as we increase selection strength. At the same time, however, we find that highly non-cooperative strategies also increase, and at a faster rate than the highly-cooperative strategies (**Extended Data Fig. 4c–e**). Combined, these effects result in overall smaller cooperation rates under strong selection.

(vi) Mutation rate μ.

In several of our figures we consider the limit of rare mutations[83–85] (e.g, **Fig. 3**,**Fig. 5a–d**). That is, we assume the mutation rate is vanishingly small, such that at any given point in time, the population employs at most two different strategies. This assumption comes with several mathematical and computational advantages[84] and it has been employed in many studies of reciprocity[23,31,69,79] and beyond[92–94]. At the same time, such an assumption entails the risk that important mixed equilibria are overlooked, because they cannot be reached by the evolutionary process. To explore the robustness of our rare mutation results, we have thus performed additional simulations for positive mutation rates (e.g., **Fig. 4**, **Fig. 5e–h**). There we have observed that our rare-mutation results do not require mutation rates to be exponentially small (as assumed by the respective theory[85]). Instead, we observe a reasonable agreement between the rare-mutation limit and our simulations for positive mutation rates once μ is of the order $1/n$ or smaller.

In **Extended Data Fig. 3g,h**, we further explore the effect of mutation rates. To gain some intuition for the effect of mutation rates, consider first the limit $\mu \to 1$. In this limit, strategy abundances are again independent of the strategies' payoffs. Every strategy is therefore played with equal likelihood. Both, the average cooperation rate as well as the average abundance of indirect reciprocity simplify to 1/2. If μ is sufficiently close to one but smaller than one, we would again expect that cooperation is slightly disfavored (because *ALLD* is a best response to a uniformly random population). While these predictions apply for all scenarios, the behavior of the system for small mutation rates depends on the exact game parameters. As mutations become sufficiently rare, we expect to recover our main text results based on the method by Imhof and Nowak[83] for the limit $\mu \to 0$ (as depicted in **Fig. 5a–d**).

Our simulations in **Extended Data Fig. 3g,h** largely follow this intuition. Interestingly, however, we again observe that cooperation based on direct reciprocity is more robust than cooperation based on indirect reciprocity when mutations occur at an intermediate rate (**Fig. 4**, **Fig. 5**). As noted in the main text, this can occur because in indirect reciprocity, the payoffs of conditional cooperators increase nonlinearly in the number of other conditional cooperators. We explore this non-linearity in more detail in the next section.

5.3 An analysis of the competition between defectors and cooperators

According to the simulations, we observe a certain bias in favor of direct reciprocity when interactions are common (i.e., when $\delta \to 1$). Surprisingly, this bias exists even if indirect information is completely reliable (i.e., when $\varepsilon \to 0$). A static equilibrium perspective cannot explain this effect. After all, Theorem 2 suggests that whenever cooperation can be sustained with direct reciprocity, it can also be sustained with indirect reciprocity if the error rate is sufficiently small. In the following, we aim to provide an evolutionary argument for this phenomenon. This argument also sheds some light on why cooperative strategies of indirect reciprocity may have problems to evolve for larger mutation rates.

To this end, let us assume that players can only choose between two strategies, a cooperative strategy $(1, 1, q, \lambda)$ and *ALLD* $= (0, 0, 0, \lambda)$. We explore the competition between these two strategies for both $\lambda = 0$ and $\lambda = 1$ in the limit of no errors, $\varepsilon \to 0$.

1. **Direct reciprocity ($\lambda = 0$).**

   In a population with k cooperators and $n-k$ defectors, we can use Eq. (8) to explicitly calculate the players' payoffs as

$$\begin{aligned} \pi_C^0 &= \frac{k-1}{n-1} \cdot (b-c) - \frac{n-k}{n-1} \cdot \Big((1-\delta)+\delta q\Big) \cdot c \\ \pi_D^0 &= \frac{k}{n-1} \cdot \Big((1-\delta)+\delta q\Big) \cdot b. \end{aligned} \tag{21}$$

   Importantly, both players' payoffs are linearly increasing in the number of cooperators k (see **Extended Data Fig. 2a,c** for an illustration of π_C^0 and π_D^0 for two different values of δ). We can simplify these payoff expressions further if we define $z := k/n$ to be the fraction of conditional cooperators in the population, and let $n \to \infty$. In that case, payoffs become

$$\begin{aligned} \pi_C^0 &= (b-c) \cdot z - (1-\delta+\delta q)c \cdot (1-z) \\ \pi_D^0 &= (1-\delta+\delta q)b \cdot z. \end{aligned} \tag{22}$$

For $z \to 0$, cooperators always obtain a lower payoff than defectors. In the other limit, $z \to 1$, cooperators outperform defectors if $q < 1 - c/(\delta b)$. Under that condition, the dynamics is bistable: each strategy is favored if it is adopted by sufficiently many players. The minimum fraction of cooperators to make cooperation beneficial can be calculated as

$$z^0 = \frac{1-\delta+\delta q}{(1-q)\delta} \cdot \frac{c}{b-c}. \tag{23}$$

This critical threshold becomes arbitrarily small as the number of repetitions increases ($\delta \to 1$), and as the cooperative strategy approaches *TFT* ($q \to 0$). That is, if only there are sufficiently many repetitions, already a small minority of *TFT* players can easily invade into an *ALLD* population.

2. **Indirect reciprocity ($\lambda = 1$).**

Again, we consider k conditional cooperators with strategy $(1, 1, q, 1)$ and $n-k$ defectors with strategy $(0, 0, 0, \lambda)$. According to Eq. (54), their respective payoffs are now

$$\begin{aligned}\pi_C^1 &= \frac{k-1}{n-1}\,\frac{1+\big((n-1)q-1\big)\delta}{1+(n-2)\delta}\,\frac{1+\big(n-2+(n-k)(1-q)\big)\delta}{1+\big(n-2-(k-1)(1-q)\big)\delta}\cdot(b-c) - \frac{n-k}{n-1}\,\frac{1+\big((n-1)q-1\big)\delta}{1+(n-2)\delta}\cdot c.\\ \pi_D^1 &= \frac{k}{n-1}\cdot\frac{1+\big((n-1)q-1\big)\delta}{1+(n-2)\delta}\cdot b\end{aligned} \tag{24}$$

The defectors' payoffs still depend linearly on the cooperators' generosity parameter q and on the number of cooperators k. However, the cooperators' payoffs are nonlinear (see **Extended Data Fig. 2b,d**). If we again define $z := k/n$, and let $n \to \infty$, the above payoffs simplify to

$$\begin{aligned}\pi_C^1 &= \frac{q+q(1-q)(1-z)}{1-(1-q)z}\cdot z(b-c) - q(1-z)c\\ \pi_D^1 &= q\cdot zb\end{aligned} \tag{25}$$

We note that $\partial \pi_C^1/\partial z > 0$ and $\partial^2 \pi_C^1/\partial z^2 > 0$ for $q > 0$. That is, the payoff of cooperators increases disproportionally in the number of other cooperators. When $z \to 0$, cooperators are again outperformed by the defectors. Conversely, when $z \to 1$, cooperators succeed if $q < 1 - c/b$. In that case, there is again a bistable competition between cooperators and defectors. The critical threshold for cooperators to be favored is now

$$z^1 = \frac{c}{b(1-q)}. \tag{26}$$

In contrast to the case of direct reciprocity, we note that this critical threshold is independent of the continuation probability. In particular, even when the game is infinitely repeated, indirect reciprocity always requires a critical mass of conditional cooperators present in the population for cooperation to succeed. The critical mass required is at least c/b.

The above results highlight that direct and indirect reciprocity differ in their relative strengths.

On the one hand, indirect reciprocity leads to a faster spread of information. As a consequence, conditional cooperators are better able to restrict the payoff of defectors. In fact, we find $\pi_D^1 \leq \pi_D^0$ for all parameter values. This faster spread of information is particularly important if players only interact for a few number of rounds. For $\delta \to 1$, this comparative advantage of indirect reciprocity disappears, and $\pi_D^1 = \pi_D^0$.

On the other hand, successful cooperation in indirect reciprocity is based on synergy effects. Cooperators only obtain a high payoff if they are present in sufficient numbers. That is, $\pi_C^1 < \pi_C^0$ for sufficiently large δ. Overall, a comparison of the thresholds for direct and indirect reciprocity yields

$$z^0 < z^1 \quad \Leftrightarrow \quad \delta \geq \frac{b}{b - c + (1-q)b}. \tag{27}$$

That is, once the continuation probability is sufficiently high, direct reciprocity requires fewer cooperative mutants to undermine an *ALLD* population.

5.4 Evolutionary dynamics for probabilistic information usage

In our previous analysis of the evolutionary dynamics, players incorporate any third-party information according to a deterministic rule. They either never take such information into account ($\lambda = 0$) or they always do so ($\lambda = 1$). In the following, we consider an evolutionary process where individuals can interpolate between these two extremes.

To this end, we need to adapt the evolutionary process as described in Section 5.1 to allow for mutant strategies with intermediate values of λ. One rather immediate way to incorporate values of λ is to assume that new mutant strategies (y', p', q', λ') are uniformly drawn from the four-dimensional unit cube $[0,1]^4$. Such a mutation scheme, however, would predominantly result in mutant strategies that rarely engage in direct reciprocity. To see why, we note that in a population of size n, two players have on average $n-2$ third-party interactions each before they have a direct encounter again. As a result, even players with comparably small values of λ have quite a high probability to update a co-player's reputation state between two direct encounters.

In the following, we formalize the above intuition. To this end, consider a player with strategy (y, p, q, λ) who just had a direct interaction with some given co-player. Let $\bar{\gamma}$ denote the probability that the player does not update the co-player's reputation until the next direct interaction of the two players. This probability can be computed as follows:

$$\bar{\gamma} = \sum_{k=0}^{\infty} \left(\frac{n-2}{n-1}\right)^k \frac{1}{n-1} (1-\lambda)^k = \frac{1}{1 + (n-2)\lambda}. \tag{28}$$

In sum on the right hand side, $\big((n-2)/(n-1)\big)^k/(n-1)$ is the probability that the co-player engages in k third party interactions until the next direct interaction happens. The remaining factor $(1-\lambda)^k$ is the probability that the focal player decides not to update the co-player's reputation after each of these third-party interactions. The resulting quantity $\bar{\gamma}$ reflects how likely the player's decision in the next game with the co-player is based on their last game with each other. We thus refer to $\bar{\gamma}$ as the effective probability that a player engages in direct reciprocity. Similarly, we refer to

$$\gamma := 1-\bar{\gamma} = \frac{(n-2)\lambda}{1+(n-2)\lambda} \tag{29}$$

as the player's effective probability to engage in indirect reciprocity. In particular, for a player with $\lambda=0$ who ignores any third-party information, we obtain $\gamma=0$ as one may expect. On the other hand, if $\lambda=1$, then $\gamma_{\max} := (n-2)/(n-1)$; in this case, the only time the player's reaction is based on direct experience is when the co-player happens not to have any third-party interactions between two direct encounters (we revisit this case in **Section 6.2**).

Importantly, the effective probability to engage in indirect reciprocity according to Eq. (29) is non-linear in λ. For example, even if a player is rather unlikely to take into account any particular piece of third-party information (e.g., $\lambda = 10\%$) may have a considerable effective probability to engage in indirect reciprocity ($\gamma \approx 82.8\%$ for a moderate population of size $n=50$). In particular, for simulations in which λ is chosen uniformly between [0,1], most mutant strategies would predominantly engage in indirect reciprocity. For our simulations in **Extended Data Fig. 12**, we thus do not employ such a mutation scheme. Instead, we generate mutant strategies (y, p, q, λ) such that the resulting γ according to Eq. (29) is distributed uniformly in $[0, \gamma_{\max}]$. In this way, we ensure that a randomly generated mutant is approximately equally likely to engage in direct and in indirect reciprocity.

For the simulations, we first consider the case that the player's value of γ is fixed. We then let γ vary between $\gamma = 0$ and $\gamma = \gamma_{\max}$. For **Extended Data Fig. 12a**, we have re-run the simulations shown in **Fig. 3a** for the case that players are comparably unlikely to interact again, $\delta=1/2$. Here we observe that players are most likely to adopt cooperative strategies when γ is large (i.e., when players have a large effective probability to engage in indirect reciprocity). Conversely, for **Extended Data Fig. 12b**, we consider the limit of infinitely many pairwise interactions, $\delta \to 1$. As already suggested by **Fig. 3b**, our results here indicate that direct reciprocity is more effective in generating high cooperation rates. While these two results are obtained in the limit of rare mutations, we obtain similar qualitative results for more abundant mutations (**Extended Data Fig. 12c,d**). When $\delta=1/2$, strategies with higher γ generate more cooperation, irrespective of the considered mutation rate. On the other hand, when $\delta=1$, strategies with smaller γ are more conducive to the evolution of cooperation (which reflects our earlier results in **Fig. 4**). We note that these simulations make a similar prediction as the theoretical results in our Theorem 2. The results suggest that in general, intermediate values of λ are not more favorable to cooperation. Instead, the highest cooperation rates are achieved either for $\lambda=0$ (i.e., $\gamma=0$) or for $\lambda=1$ (i.e., $\gamma=\gamma_{\max}$).

In a second step, we then allow the value of γ to co-evolve along with the other strategy parameters

y, p, and q. Repeating the simulations done for **Fig. 5a–c**, we consider three different scenarios. (*i*) If errors are abundant and players only interact for a few rounds, we observe that no cooperation evolves and that selection among the player's γ values is neutral (**Extended Data Fig. 12e**). (*ii*) If errors are rare and players interact for an intermediate number of rounds, cooperation can evolve (**Extended Data Fig. 12f**). In that case, we also observe that players tend to have a higher effective probability to engage in indirect reciprocity. In particular, among all possible values of γ, players most abundantly choose $\gamma \approx \gamma_{\text{max}}$. (*iii*) If errors occur at an intermediate rate and players interact often, players show a bias towards sustaining cooperation based on direct reciprocity (**Extended Data Fig. 12g**). Now the most abundant strategies adopt $\gamma \approx 0$.

In a last step, we have systematically varied the considered error rate and the continuation probability (**Extended Data Fig. 12h**). Compared to the baseline case in **Fig. 5d**, we observe some notable differences. On the one hand, when errors are abundant, players seem to have more difficulties to establish cooperation. On the other hand, also the average value of γ seems to be closer to $1/2$ (note that the color legend in **Extended Data Fig. 12h** is rescaled to allow for a better contrast between the different regions). Nevertheless, the general patterns outlined in **Fig. 5d** are conserved. Cooperation is most likely to evolve when the continuation probability is large and when errors are rare. At the same time, players show a bias towards indirect reciprocity if they only interact for intermediately many rounds, whereas they prefer direct reciprocity when there are many rounds and many errors.

6 Model extensions

6.1 The impact of errors and incomplete information

Our baseline model introduced in Section 2 is based on a number of simplifying assumptions. In particular, we assumed that only third-party interactions are subject to perception errors, that player always implement the correct action, that both actions (C and D) are equally likely to be misperceived, and that all players can observe all interactions in a population. In the following, we discuss the impact of each of these assumptions separately.

Direct reciprocity with perception errors. We first discuss how our results need to be adapted when players misperceive the actions of a direct interaction partner with probability $\varepsilon_0 > 0$. For better clarity, we now refer the probability that players misperceive third-party interactions as ε_1. The baseline model corresponds to the case $\varepsilon_0 = 0$ and $\varepsilon_1 = \varepsilon$. In general, it is reasonable to assume that indirect observations are more prone to misinterpretation than direct observations, in which case $\varepsilon_0 \leq \varepsilon_1$. When direct reciprocity is subject to perception errors, the main equation (1) becomes

$$\begin{aligned} x_{ij}(t+1) &= (1-\bar{w})\, x_{ij}(t) \\ &+ w\Big(\big((1-\varepsilon_0)x_{ji}(t)+\varepsilon_0(1-x_{ji}(t))\big)\, p_i + \big((1-\varepsilon_0)(1-x_{ji}(t)) + \varepsilon_0 x_{ji}(t)\big)\, q_i\Big) \\ &+ (\bar{w}-w)\,(1-\lambda_i)\, x_{ij}(t) \\ &+ \lambda_i\, w \sum_{l\neq i} \big[\big((1-\varepsilon_1)x_{jl}(t) + \varepsilon_1(1-x_{jl}(t))\big)\, p_i + \big((1-\varepsilon_1)(1-x_{jl}(t)) + \varepsilon_1 x_{jl}(t)\big)\, q_i\big]. \end{aligned} \tag{30}$$

We note that the new parameter ε_0 only affects the second line of this recursion, which captures the direct interaction between i and j (**Extended Data Fig. 1f**). As in Section 2, we can use this equation to calculate the players' average cooperation rates and payoffs. Again, we are interested in which strategies are able to sustain cooperation. To this end, we need to slightly modify our definition of cooperative Nash equilibria. We say a strategy $\sigma=(y,p,q,\lambda)$ is a cooperative Nash equilibrium if

(i) σ is a Nash equilibrium, and

(ii) the cooperation rate in a homogeneous σ-population approaches one as both $\varepsilon_0\to 0$ and $\varepsilon_1\to 0$.

Based on recursion (30), the equilibrium results of Section 4 can be adapted as follows.

Theorem 4 (Cooperative Nash equilibria when direct interactions are subject to perception errors)**.**

1. *A strategy $\sigma=(y,p,q,\lambda)$ is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\lambda, \quad \text{with} \quad r^*_\lambda=\frac{1+(n-2)\delta\lambda}{(1-2\varepsilon_0)+(n-2)(1-2\varepsilon_1)\lambda}\cdot\frac{c}{\delta b}. \tag{31}$$

2. *In particular, cooperation can be sustained through direct reciprocity ($\lambda=0$) if and only if $\delta\geq\delta_0$ with*

$$\delta_0 = \frac{1}{1-2\varepsilon_0}\cdot\frac{c}{b}. \tag{32}$$

It can be sustained through indirect reciprocity ($\lambda=1$) if and only if $\delta\geq\delta_1\in[0,1]$ where

$$\delta_1=\frac{c}{(1-2\varepsilon_0)\,b+(n-2)\big((1-2\varepsilon_1)\,b-c\big)}. \tag{33}$$

In **Fig. 2d**, we illustrate the equilibrium conditions (32) and (33) for the special case $\varepsilon_0=\varepsilon_1$. As one may expect, it is more difficult to sustain cooperation when also direct interactions are subject to perception errors. This increased difficulty is reflected in the additional $(1-2\varepsilon_0)$-terms in the denominator of equations (32) and (33). However, the impact of ε_0 on the threshold value δ_1 for indirect reciprocity vanishes as the population size n becomes large. Again, this is intuitive: if a players' reputation is based on his entire behavior, errors that only affect direct encounters become negligible in large populations.

To explore how perception errors in direct interactions affect the evolutionary dynamics of strategies, we have re-run the simulations in **Fig. 5d**, but now assuming that $\varepsilon_0=\varepsilon_1=\varepsilon$. As shown in **Extended Data Fig. 5b**, there are two effects. First, players are now less likely to cooperate, especially for high error rates. Second, players become more likely to take indirect information account (as it is now just as reliable as direct information).

Implementation errors. In addition to perception errors, models of direct and indirect reciprocity often consider an alternative source of noise in the form of implementation errors or trembling hand errors [95,96]. To account for these kinds of errors, we assume that when a player intends to cooperate (defect), he mistakenly defects (cooperates) with probability e. Under this assumption, Eq. (1) for player i's probability to be in the good state with respect to co-player j becomes

$$\begin{aligned}
x_{ij}(t{+}1) &= (1{-}\bar{w})\, x_{ij}(t) \\
&\quad + w\Big(\big((1{-}e)x_{ji}(t){+}e(1{-}x_{ji}(t))\big)\, p_i + \big((1{-}e)(1{-}x_{ji}(t)){+}ex_{ji}(t)\big)\, q_i\Big) \\
&\quad + (\bar{w}{-}w)\,(1-\lambda_i)\, x_{ij}(t) \\
&\quad + \lambda_i\, w \sum_{l\neq i} \Big[\Big((1{-}e)(1{-}\varepsilon)x_{jl}(t){+}(1{-}e)\varepsilon\big(1{-}x_{jl}(t)\big){+}e(1{-}\varepsilon)\big(1{-}x_{jl}(t)\big){+}e\varepsilon x_{jl}(t)\Big)\, p_i \\
&\qquad + \Big(e(1{-}\varepsilon)x_{jl}(t){+}e\varepsilon\big(1{-}x_{jl}(t)\big){+}(1{-}e)(1{-}\varepsilon)\big(1{-}x_{jl}(t)\big){+}(1{-}e)\varepsilon x_{jl}(t)\Big)\, q_i\Big].
\end{aligned} \tag{34}$$

Here, the second and fourth line have changed in comparison to Eq. (1). For example, the second line reflects that in direct interactions, we need to distinguish four cases, depending on whether or not the co-player intended to cooperate, and whether or not there was an implementation error. Also the payoff formula (8) needs to be adapted accordingly,

$$\pi_i = \frac{1}{n{-}1} \sum_{j\neq i} \big((1{-}e)x_{ji}{+}e(1{-}x_{ji})\big)\, b - \big((1{-}e)x_{ij} + e(1{-}x_{ij})\big)\, c. \tag{35}$$

Again, we are interested in those Nash equilibria in which everyone is fully cooperative as errors become rare (that is, when $\varepsilon \to 0$ and $e \to 0$).

Theorem 5 (Cooperative Nash equilibria under implementation errors)**.**

1. *A strategy $\sigma = (y, p, q, \lambda)$ is a cooperative Nash equilibrium if and only if*

$$y{=}1, \quad p{=}1, \quad q{=}1{-}r^*_\lambda, \quad \textit{with} \quad r^*_\lambda = \frac{1 + (n{-}2)\delta\lambda}{(1{-}2e)\big(1{+}(n{-}2)(1{-}2\varepsilon)\lambda\big)} \cdot \frac{c}{\delta b}. \tag{36}$$

2. *In particular, cooperation can be sustained through direct reciprocity if and only if $\delta \geq \delta_0$ with*

$$\delta_0 = \frac{1}{1{-}2e} \cdot \frac{c}{b}. \tag{37}$$

It can be sustained through indirect reciprocity if and only if $\delta \geq \delta_1 \in [0,1]$ where

$$\delta_1 = \frac{c}{(1{-}2e)\, b + (n{-}2)\big((1{-}2e)(1{-}2\varepsilon)\, b - c\big)}. \tag{38}$$

Interestingly, when we compare Theorem 5 with the respective conditions in Section 4, we note that the effect of implementation errors is equivalent to a rescaling of the benefit parameter from b to $(1-2e)b$. That is, with respect to the static equilibrium predictions, increasing the rate of implementation errors by 10% has the same effect as reducing the benefit of the game by 20%. To explore the effect of implementation errors on the evolution of cooperation, we have re-run the simulations in **Fig. 5d**, but now for $e = 0.01$ instead of $e = 0$ (**Extended Data Fig. 5c**). As one may expect, the addition of implementation errors renders the evolution of cooperation more difficult. In the new scenario, players need to interact for a larger number of rounds to establish substantial cooperation rates.

Asymmetric errors. One way how misperceptions can occur is when the acting player herself has an incentive to misrepresent her action. This is particularly relevant for defectors, who may wish to conceal the true nature of their actions. Excellent models of strategic miscommunication in indirect reciprocity can be found in the papers by Nakamaru and Kawata and by Seki and Nakamaru[73,74]. In the context of our model we can approximate the workings of strategic miscommunication by assuming that the two possible actions C and D have different probabilities to be misperceived, ε_C and ε_D. A scenario where defectors are more likely to misrepresent their actions can then be represented by assuming $\varepsilon_D > \varepsilon_C$. For asymmetric errors, we obtain the following recursion for the pairwise probability to assign a good reputation to a co-player,

$$\begin{aligned} x_{ij}(t{+}1) = (1{-}\bar{w})\, x_{ij}(t) & \\ + w\left(x_{ji}(t)\, p_i + (1{-}x_{ji}(t))\, q_i\right) & \\ + (\bar{w}{-}w)\,(1{-}\,\lambda_i)\, x_{ij}(t) & \\ + \lambda_i \sum_{l\neq i} w\left[\left((1{-}\varepsilon_C)x_{jl}(t) + \varepsilon_D(1{-}x_{jl}(t))\right) p_i + \left((1{-}\varepsilon_D)(1{-}x_{jl}(t)) + \varepsilon_C x_{jl}(t)\right) q_i\right]. & \end{aligned} \tag{39}$$

The players' payoffs are again defined by Eq. (8). In the case of asymmetric errors, we speak of a strategy as a cooperative Nash equilibrium if it is stable and if the cooperation rate against itself approaches one if both $\varepsilon_C \to 0$ and $\varepsilon_D \to 0$. We obtain the following characterization:

Theorem 6 (Cooperative Nash equilibria under asymmetric errors)**.**

1. *A strategy $\sigma = (y, p, q, \lambda)$ is a cooperative Nash equilibrium if and only if*

$$y{=}1, \quad p{=}1, \quad q{=}1{-}r^*_\lambda, \quad \textit{with} \quad r^* = \frac{(1+(n{-}2)\delta\lambda)}{1+(n{-}2)\,(1{-}\varepsilon_D{-}\varepsilon_C)\lambda}\frac{c}{\delta b}. \tag{40}$$

2. *In particular, cooperation can be sustained through direct reciprocity if and only if $\delta \geq \delta_0$ with*

$$\delta_0 = \frac{c}{b}. \tag{41}$$

It can be sustained through indirect reciprocity if and only if $\delta \geq \delta_1 \in [0,1]$ where

$$\delta_1 = \frac{c}{b + (n{-}2)\left((1{-}\varepsilon_D{-}\varepsilon_C)b{-}c\right)} \tag{42}$$

In case $\varepsilon_C = \varepsilon_D$, these conditions recover the respective results of the baseline model, which is reassuring. In **Extended Data Fig. 5d**, we show simulation results for the asymmetric case where $\varepsilon_C = 0$ and $\varepsilon_D = \varepsilon > 0$. For the parameter range considered previously, $0 \le \delta \le 1$ and $0.001 \le \varepsilon \le 0.1$, we observe more cooperation compared to our baseline model. This occurs because the effective rate at which errors happen is now reduced, because only some of the players' actions are subject to noise.

However, we note that in contrast to the previous scenarios, it is not unreasonable to assume that error probabilities may exceed 10% when they are the result of strategic miscommunication. Larger values of ε_D can impede cooperation considerably. In fact, even if $\varepsilon_C = 0$, it follows from Eq. (42) that indirect reciprocity is no longer able to sustain cooperation if $\varepsilon_D > (1-c/b)(n-1)/(n-2)$. That is, for cooperation to evolve through indirect reciprocity, there need to be limits on how well defectors can deceive others.

Incomplete information. In our baseline model we assume that each player is informed about everyone else's interactions. This assumption of complete information though standard is unrealistic, especially if the population is large. In the following, we thus introduce an additional parameter $\nu \in [0,1]$. This parameter reflects which fraction of third-party interactions players observe on average. With incomplete information, the main equation (1) now becomes

$$\begin{aligned} x_{ij}(t+1) &= (1-\bar{w})\, x_{ij}(t) \\ &\quad + w\,\big(x_{ji}(t)\, p_i + (1-x_{ji}(t))\, q_i\big) \\ &\quad + (\bar{w}-w)\,(1-\nu\lambda_i)\, x_{ij}(t) \\ &\quad + w\nu\lambda_i \sum_{l\neq i,j} \Big(\big(1-\varepsilon\big)x_{jl}(t) + \varepsilon\big(1-x_{jl}(t)\big)\Big)\, p_i + \Big(\big(1-\varepsilon\big)\big(1-x_{jl}(t)\big) + \varepsilon x_{jl}(t)\Big)\, q_i. \end{aligned} \tag{43}$$

Here, the third and fourth line have changed in comparison to Eq. (1). That is, a player's probability λ_i to take into account indirect information is now scaled by a factor of ν, the probability that this information is learned in the first place. A player with $\lambda_i = 1$ no longer takes every action of the other group members into account when assigning reputations, but every action he is aware of. The condition for cooperation to be feasible is now given as follows.

Theorem 7 (Cooperative Nash equilibria under incomplete information)**.**

1. *A strategy $\sigma = (y, p, q, \lambda)$ is a cooperative Nash equilibrium if and only if*

$$y = 1, \quad p = 1, \quad q = 1 - r^*_\lambda, \quad \textit{with} \quad r^*_\lambda = \frac{1 + (n-2)\delta\nu\lambda}{1 + (n-2)(1-2\varepsilon)\nu\lambda} \cdot \frac{c}{\delta b}. \tag{44}$$

2. *The condition for cooperation under direct reciprocity is unchanged, $\delta \ge \delta_0 = c/b$. For indirect reciprocity there exists a cooperative Nash equilibrium if and only if $\delta \ge \delta_1 \in [0,1]$ where*

$$\delta_1 = \frac{c}{b + (n-2)\big((1-2\varepsilon)\, b - c\big)\nu}. \tag{45}$$

In the limiting case that players obtain no third-party information at all, such that $\nu = 0$, the two threshold values coincide, $\delta_0 = \delta_1$, as expected. In **Extended Data Fig. 5d**, we present evolutionary simulation results when there is almost no third-party information ($\nu = 0.01$, such that only every 100th interaction is observed on average). In that case, players who take indirect information into account are almost indistinguishable from those players who do not. Indeed, we find that cooperation only evolves once it can be sustained through direct reciprocity, and that indirect reciprocity is generally disfavored.

6.2 A model of pure indirect reciprocity

Our baseline model makes two assumptions on how individuals assign reputations to each other: (*i*) Individuals always update a co-player's reputation after a direct interaction, and (*ii*) if the respective co-player interacts with a third party, the co-player's reputation is updated with probability λ. These assumptions imply that a player with $\lambda = 0$ acts based on direct experiences only, whereas a player with $\lambda = 1$ equally takes into account all of a co-player's interactions, no matter whether the focal player was directly involved. Since interactions occur randomly, however, it may happen that the same two players are chosen to interact for two consecutive rounds. In such a case, the players' decisions in the second round are necessarily based on their respective direct experience made in the first round – even if players generally base their decisions on indirect reciprocity (that is, even if $\lambda = 1$). On a conceptual level, this effect may make it more difficult to compare the two modes of reciprocity. It can thus be useful to study a model where individuals use pure versions of each behavior (such that individuals who opt for indirect reciprocity neglect any additional direct information they may have). In the following we present such a model, and we show that it leads to qualitatively similar conclusions.

For this model, we assume the players' strategies are given by a 4-tuple (y, p, q, κ). The interpretation of the first three entries is the same as in the baseline model. For the last entry, we assume that if a focal player directly directly interacts with a given co-player, the co-player's reputation is updated with probability $1 - \kappa$. In contrast, if the co-player interacts with a third-party, the co-player's reputation is updated with probability κ. This model implies that for $\kappa = 0$ only direct interactions lead to a reputation update, whereas for $\kappa = 1$ reputations are only updated after third-party interactions. Similar to Eq. (1) for the baseline model, we can derive a recursion for the probability $x_{ij}(t)$ that player i considers j to be good at time t. This recursion now takes the form

$$\begin{aligned} x_{ij}(t+1) = {} & (1-\bar{w})\, x_{ij}(t) \\ & + w(1-\kappa_i)\big(x_{ji}(t)\, p_i + (1-x_{ji}(t))\, q_i\big) \\ & + w\kappa_i \cdot x_{ij}(t) \\ & + (\bar{w}-w)\,(1-\kappa_i)\, x_{ij}(t) \\ & + w\kappa_i \sum_{l\neq i,j} \Big((1-\varepsilon)x_{jl}(t) + \varepsilon\big(1-x_{jl}(t)\big)\Big)\, p_i + \Big((1-\varepsilon)\big(1-x_{jl}(t)\big) + \varepsilon x_{jl}(t)\Big)\, q_i. \end{aligned} \tag{46}$$

This recursion only differs from Eq. (1) in the second and third line. These two lines both refer to direct interactions of players i and j, which happen with probability w. For the model considered here, direct interactions lead to a reputation update with probability $1-\kappa_i$. In that case, the new reputation depends on co-player j's action, as encoded by $x_{ji}(t)$, and on the values of p_i and q_i. Otherwise, with probability κ_i, player j's reputation remains unaffected.

Based on this recursion (46) can now derive the corresponding versions of all results of the baseline model. We summarize the respective findings in the following.

Theorem 8 (Cooperative Nash equilibria for the model with pure indirect reciprocity)**.**
Consider a group of size $n \geq 3$.

1. *A strategy $\sigma = (y, p, q, \kappa)$ is a cooperative Nash equilibrium if and only if*

$$y=1, \quad p=1, \quad q=1-r^*_\kappa, \quad \textit{with} \quad r^*_\kappa = \frac{1+(n-3)\delta\kappa}{1-\kappa+(n-2)(1-2\varepsilon)\kappa}\,\frac{c}{\delta b}. \tag{47}$$

2. *In particular, if players only use direct reciprocity ($\kappa = 0$), we obtain the same condition for the existence of a cooperative Nash equilibrium as in the baseline model, $\delta \geq \delta_0 = c/b$. For indirect reciprocity ($\kappa = 1$) there is a cooperative Nash equilibrium if and only if $\delta \geq \delta_1 \in [0,1]$ where*

$$\delta_1 = \frac{c}{b(n-2)(1-2\varepsilon) - c(n-3)}. \tag{48}$$

For any $\kappa > 0$, we note that the equilibrium value of q according to Eq. (47) approaches $q = 1 - c/(1\text{–}2\varepsilon)/b$ in the limit of large populations. This limit coincides with the respective limit of the baseline model according to Eq. (16). This is unsurprising: when populations are sufficiently large, the probability that two particular players interact with each other twice in a row becomes negligible. In that case, already our baseline model yields a model of pure indirect reciprocity. But even for finite population sizes, the two models typically yield similar predictions. For example, for the parameters in **Fig. 3** ($b/c = 5$, $n = 50$, $\varepsilon = 0$) the baseline model suggests that full cooperation is feasible for $\delta_0 = 0.2$ (when $\lambda = 0$) and $\delta_1 = 0.0051$ (when $\lambda = 1$). For the alternative model considered here, the respective values become $\delta_0 = 0.2$ (when $\kappa = 0$) and $\delta_1 = 0.0052$ (when $\kappa = 1$), respectively. As a result, we also observe a similar dynamics in evolutionary simulations (see **Extended Data Fig. 11**).

6.3 Finite-state automata with multiple states

So far we have studied the evolution of direct and indirect reciprocity in the most simple strategy space possible. Players can only assign two possible reputations to their co-players, and a player's updated reputation only depends on the last action taken into account. In the following two subsections, we sketch how our framework can be extended to scenarios that do not meet these two assumptions.

Strategies represented by finite-state automata with multiple states. We begin by allowing players to assign more nuanced reputations to their co-players. To this end, we consider players who use finite state automata with more than two states. In addition to the dichotomous characterization of a player being either 'good' or 'bad', this allows us to capture models in which players can have a third, 'unknown' or 'neutral' reputation. Such third states have been important in models in which players can forget[76]. Similarly, these automata also allow us to capture scenarios in which it takes multiple defections in memory to yield a bad reputation[66,67].

For this model extension, we assume that each player's strategy takes the form $(\Omega, \gamma, \tau_0, \tau, \lambda)$. Here, the first component $\Omega = \{\omega_1, \ldots, \omega_m\}$ is the number of (reputational) states the player does distinguish. The second component $\gamma\colon \Omega \to \{C, D\}$ determines for each assigned reputational state whether or not to cooperate with a co-player who has the respective reputation. The third component $\tau_0 \in \Delta^\Omega$ determines the initial reputational state of the co-players. Here, $\Delta^\Omega = \{x \in [0,1]^m \mid x_1 + \ldots + x_n = 1\}$ is the set of probability distributions over the set of states. The fourth component $\tau : \Omega \times \{C, D\} \to \Delta^\Omega$ is the transition function. It determines how likely the updated reputation of a given co-player is $\omega' \in \Omega$, depending on the previously assigned reputation $\omega \in \Omega$ and on the co-player's action $a \in \{C, D\}$. As before, we assume that the co-player's reputation is always updated after a direct interaction. Whether or not the co-player's reputation is also updated if the co-player interacts with a third party depends on the player's receptivity λ. Third party interactions are ignored when $\lambda = 0$, whereas they are fully included when $\lambda = 1$. In the first case, we again obtain a model with direct reciprocity only. In that case, our framework captures previous studies on repeated games among players with finite state automata[17,18,26]. If $\lambda = 1$, we obtain a model of indirect reciprocity. In that case, we can capture previous models in which players can have more than two possible reputations[49,75,76].

The reactive strategies considered in the previous sections represent a special case of these finite state automaton strategies. For a reactive strategy $\sigma = (y, p, q, \lambda)$ we have

$$\Omega = \{G, B\}, \quad \gamma(\omega) = \begin{cases} C & \text{if } \omega = G \\ D & \text{if } \omega = B \end{cases}, \quad \tau_0 = (y, 1-y), \quad \tau(\omega, a) = \begin{cases} (p, 1-p) & \text{if } a = C \\ (q, 1-q) & \text{if } a = D. \end{cases} \tag{49}$$

From this representation, the two important simplifications of reactive strategies become immediately apparent: (*i*) reputations are binary, and (*ii*) when a co-player's reputation is updated, the new reputation only depends on the co-player's action but not on her previous reputation. This latter assumption has been crucial for the explicit calculation of the players' payoffs in Section 3. Due to this assumption, the pairwise reputation variable $x_{ij}(t+1)$ can be written as a linear function of the respective quantities $x_{kl}(t)$ in the previous round. This no longer needs to hold for more general finite state automata, where transitions depend on both the current state and the respective co-player's action. Instead of calculating the payoffs explicitly, we instead use numerical simulations of the game dynamics to obtain the players' payoffs in the following.

Three examples of finite-state automata with multiple states. We illustrate this approach with three examples (**Extended Data Fig. 6**). In each case, we assume the respective finite state automaton has three states, 'good' (G), 'neutral' (N), and 'bad' (B). Initially, players assign a good reputation to everyone. Each time the co-player defects, the co-player's reputation deteriorates (from good to neutral, or from neutral to bad, respectively). Similarly, each time the co-player cooperates, his reputation improves. Players are assumed to cooperate against good opponents and to defect against bad opponents. The three automata differ in how they deal with opponents with a neutral reputation: we consider the case that players may either cooperate ($A1$, **Extended Data Fig. 6a**), cooperate with 50% probability ($A2$, **Extended Data Fig. 6b**), or defect ($A3$, **Extended Data Fig. 6c**). The automata $A1$ and $A3$ have been referred to as 'generous discriminator' and 'rigorous discriminator', respectively[75].

Stability of the three-state automata against ALLD and ALLC. In a next step we considered a scenario in which $n-1$ residents employ the respective automaton strategy. The remaining player either employs *ALLD* or *ALLC*. Keeping the population composition fixed, we simulate the players' payoffs for various values of $\lambda \in [0, 1]$ (**Extended Data Fig. 6d–f**). We used the fixed game parameters $b = 5$, $c = 1$, $n = 50$, and $\varepsilon = 0.05$. To compute the payoffs, we use the same simulation scheme as in a previous study[69]. That is, initially, all players are assumed to assign a good reputation to each other. Then we simulate a game with $2{\cdot}10^6$ rounds. To compute the players' payoffs , we average over the second half of these rounds; in this way, we avoid any transient effect during the early rounds arising from our assumption on the initial reputation assignments. We checked that averaging over all rounds would not alter our conclusions.

For direct reciprocity only ($\lambda = 0$), we observe that all of the three automaton strategies are stable with respect to invasion by *ALLD*, and neutrally stable with respect to invasion by *ALLC*. However, once indirect information is considered ($\lambda > 0$), $A2$ and $A3$ can both be invaded by *ALLC*. Moreover, if residents adopt $A3$, they fail to cooperate with each other altogether. Only the first automaton $A1$ is stable against *ALLC* and *ALLD* for all considered λ values. Automaton $A1$ can be considered as a threshold strategy in the sense of Berger and Grüne[67]: players are certain to defect only if the co-player defected twice in a row. For appropriate parameter values, such threshold strategies can sustain cooperation although they only rely on first-order information[66,67]. Our results in **Extended Data Fig. 6d** suggest that the same is true when players blend direct and indirect reciprocity (i.e., for $0 < \lambda < 1$).

Evolutionary competition between the three-state automata, ALLD, and ALLC. In a final step, we sketch how our framework can be used to explore the evolutionary performance of finite-state automata which depend on both, direct and indirect information. In the following, we assume players use a fixed receptivity $\lambda = 0.1$ (but it should become clear how our methods extend to more general settings). Players can choose between three different strategies: *ALLD*, *ALLC*, and one of the three automata discussed above.

In contrast to reactive strategies, there is no known efficient payoff formula for arbitrary finite-state automata in the context of indirect reciprocity with private information. We therefore determined the players' payoffs by simulation. To this end, we considered all possible population compositions (k_A, k_C, k_D). Here, k_A is the number of players who use the respective finite-state automaton strategy, k_C is the number of unconditional cooperators, and k_D is the number of defectors. The possible population compositions are those for which $k_A + k_C + k_D = n$. Based on these payoffs, we again employ the process described in Section 5.1.

First, we considered the dynamics for $\beta = 1$ in the limit of rare mutations, $\mu \to 0$, for each of the three finite-state automata. For the given parameter values, we find that only the first automaton A_1 is played with substantial frequency (**Extended Data Fig. 7a–c**). We observe that rare defectors are disfavored to invade into an A_1 population. However, cooperators can invade through (almost) neutral drift, because they get a payoff only slightly lower than the residents'. Because unconditional cooperators are in turn easily invaded by *ALLD*, there is a cyclic dynamics. Most of the time, the population either adopts A_1 (65.4%) or *ALLD* (32.8%). As a result, the average cooperation rate over time is below 70%. The other two finite-state automata generate even lower cooperation rates; A_2 is quickly invaded by *ALLC*, whereas A_3 can be invaded by both, *ALLC* and *ALLD*.

Next we have explored the dynamics for a mutation rate of $\mu = 0.01$, such that different strategies may coexist in a population. For all three automata we find that most of the time, the population either is in the vicinity of *ALLD*, or in the vicinity of the edge between the the finite-state automaton and *ALLC* (**Extended Data Fig. 7d–f**). How often these two neighborhoods are visited depends on the considered automaton strategy. For A_1 and A_2, we observe substantial cooperation, whereas for A_3 players adopt *ALLD* most of the time.

In a final step we have re-run these calculations for different parameter values. We have varied the benefit of cooperation, the selection strength, and the mutation rate. As baseline parameters we took the values used in (**Extended Data Fig. 7d–f**). For all parameter values, we find that A_1 is most likely to induce cooperation (**Extended Data Fig. 7g–i**). However, also the second automaton can yield substantial cooperation rates.

Most strikingly, while the mutation rate had a negative impact on cooperation for reactive strategies (**Fig. 4**), here we find that an intermediate mutation rate is most favorable to cooperation. However, in our view the two sets of simulations need to be compared with caution. In **Fig. 4**, we have considered the entire space of all strategies of a given complexity. In particular, the space is balanced: for every conditional strategy that becomes more cooperative if the co-player cooperates, there is an analogous strategy that reduces its cooperation rate. Only under this assumption, an increase of mutation rates does not per se change the resulting cooperation rate. In contrast, the simulations in this section are based on only a small sample of possible strategies, and the strategy space is not balanced. This shortcoming is not specific to our **Extended Data Fig. 7**, but it is rather common in the indirect reciprocity literature[49–52,55–58,69]. One reason for this prevalence of unbalanced models is the difficulty of deriving the players' payoffs. In the absence of a general payoff equation that applies to all strategies of a given complexity, researchers have to focus on strategies that seem most relevant. Such an approach does not naturally induce balanced strategy sets. In this sense, we consider the simulations performed for our baseline model as more transparent. They allow for all strategies of some given complexity, without any preselection on the part of the researcher.

6.4 Higher-order strategies

Another class of strategies that has received considerable attention in the literature on indirect reciprocity (but less so in direct reciprocity) is the class of higher-order social norms[53,60,61]. The first-order social norms correspond to the reactive strategies considered in the baseline model. When two players interact, then the updated reputation of each player only depends on whether or not that player has cooperated. Second-order norms additionally take the reputation of the respective opponent into account. An example of such a second-order norm is stern judging[55,59,79]. This norm suggests that to maintain a good reputation, one should cooperate with good opponents and defect against bad opponents. Finally, in third-order norms the updated reputation of some player additionally depends on the player's previous reputation. In that way, higher-order social norms take an increasing number of information into account when assigning a reputation to another group member.

Using an exhaustive method, Ohtsuki and Iwasa have shown that under public information, there are exactly eight deterministic third-order social norms that can maintain cooperation[53,54]. These norms are called the leading-eight. They consist of two components, an assessment-rule and an action rule (see **Extended Data Fig. 8a**). The assessment rule determines how to update the reputation of other group members, depending on their actions, their previous reputation, and the reputation of the opponent. The action rule determines whether to cooperate; this decision may depend on one own's reputation, as well as on the reputation of the opponent. Crucially, Ohtsuki and Iwasa assume public information. In the following, we explore the performance of the leading-eight rules when reputations are based on both direct interactions and indirect observations, and when information is private.

Representation of third-order strategies. To this end, we now assume the players' strategies have the form $\sigma_i = (\alpha_i, \beta_i, \lambda_i)$. The first component is player i's assessment rule,

$$\alpha = (\alpha_{GCG},\ \alpha_{GCB},\ \alpha_{BCG},\ \alpha_{BCB},\ \alpha_{GDG},\ \alpha_{GDB},\ \alpha_{BDG},\ \alpha_{BDB}) \in \{G, B\}^8. \tag{50}$$

For example, an entry of $\alpha_{GCB} = G$ means that i deems it as good if a good population member j cooperates with a bad population member k. Importantly, whether or not j and k had a good and bad reputation to start with, depends on the individual perspective of player i (that is, it depends on $x_{ij}(t)$ and $x_{ik}(t)$ at time t). The second component of player i's strategy corresponds to his action rule,

$$\beta = (\beta_{GG},\ \beta_{GB},\ \beta_{BG},\ \beta_{BB}) \in \{C, D\}^4. \tag{51}$$

For example, if player i considers himself to be in a good standing but his co-player to be bad, then an action rule with $\beta_{GB} = D$ would prescribe to defect. Again, these assessments need to be taken from i's individual perspective. To this end, we assume that each player i now also has one additional automaton that captures i's own reputation. For this automaton, we set $x_{ii}(t) = 0$ if player i considers himself to be bad at time t (which depends on i's assessment rule and his previous interactions). Otherwise, if $x_{ii}(t) = 1$, player i deems himself as good. We assume the automaton that captures player i's self-perception is updated every time player i interacts with a co-player. With regard to the reputation of others, we assume that i updates a co-player j's reputation after a direct interaction with probability 1, and after an indirect observation with probability λ_i, as before.

Stability of the leading-eight strategies against ALLD and ALLC. To explore the stability of third-order social norms in the presence of both direct and indirect reciprocity, we proceed along the same lines as in the previous section on finite-state automata. Again, we first consider a population in which $n-1$ players adopt some leading-eight strategy L_i. The remaining player either adopts *ALLC* or *ALLD*. To compute the player's payoffs for various values of λ, again we simulate the game dynamics for fixed game parameters $b=5$, $c=1$, $n=50$, and $\varepsilon=0.05$. Our results for $\lambda>0$ resemble previous findings on indirect reciprocity[69]: in the presence of perception errors, all leading-eight strategies are susceptible to invasion. Either a single *ALLC* player or a single *ALLD* player obtains a higher payoff than the residents (**Extended Data Fig. 8b–i**). Only for $\lambda=0$ (when perception errors are absent), the leading-eight strategies are stable against both mutant strategies.

Evolutionary competition between the leading-eight, ALLD, and ALLC. In a next step, we have again explored how the leading-eight perform in an evolutionary context, when competing with the two unconditional strategies. We use exactly the same setup as in our previous analysis on finite-state automata. That is, we fixed a receptivity value $\lambda = 0.1$ for all population members. Then we pre-computed the payoffs with simulations for all possible population compositions (k_L, k_C, k_D) where $k_L+k_C+k_D=n$ (using the above payoff parameters). Based on these payoffs, we first considered the rare-mutation dynamics. As in a previous analysis for indirect reciprocity only[69], we find that all of the leading-eight strategies have problems to persist in the population (**Extended Data Fig. 9a–h**). The only strategy that achieves notable frequencies in the population is L_8. However, L_8 is sensitive to noise, and thus tends to defect in the presence of errors.

In a next step, we have considered the evolutionary process for a positive mutation rate, $\mu=0.01$. As we observed for the finite-state automaton strategies, populations tend to be clustered in one of two regions of the space of all population compositions (**Extended Data Fig. 9i–p**). Either most players adopt *ALLD*, or the population consists of a mixture of leading-eight players and unconditional cooperators. Cooperative strategies are most abundant for three of the eight cases, for L_1, L_2, and L_7, as observed previously in the case of indirect reciprocity only[69].

Finally, we have again explored how often players cooperate on average as we change three key parameters, the benefit of cooperation, the strength of selection, and the mutation rate. Across all parameters considered, we find that only the three previously identified strategies L_1, L_2, and L_7 can result in a population that predominantly cooperates (**Extended Data Fig. 9q–s**). For these strategies we find more cooperation when the mutation rate is intermediate, for $0.01 \leq \mu \leq 0.1$. This again disagrees with the findings of our baseline model, according to which such mutation rates are detrimental to indirect reciprocity. However, as explained before, the two scenarios should not be compared directly.

The simulations for our baseline model consider a complete and balanced strategy space. In contrast, the simulations **Extended Data Fig. 9** only allow for three particular strategies, none of which being anti-reciprocal (i.e., none of which reducing its cooperation rate in response to a cooperative co-player). When a strategy space is not balanced, larger mutation rates may yield more cooperation even in the absence of selection.

To address such problems, it would be desirable to have simulations in which all third-order strategies (or strategies with different values of λ) compete. However, because there is no general payoff formula for third-order strategies under private information, such an approach is computationally out of reach. For our baseline model, we have thus analyzed a more elementary strategy space, the set of all (stochastic) first-order strategies. We find that even in this simpler strategy space, there is an unexpected strategy of indirect reciprocity that can maintain cooperation under appropriate conditions. This strategy is *GSCO* (Generous Scoring). *GSCO* does not require higher order information. All it requires is that players sometimes forgive when they witness a defection.

7 Appendix

7.1 Efficient computation of payoffs

Here we show how the payoffs according to Eq. (8) can be computed more efficiently by taking into account that players with the same strategy receive the same expected payoff.

To this end, suppose the population contains players with s different strategies in total. Let k_i denote the number of players using strategy i, such that $\sum_{i=1}^{s} k_i = n$. Slightly abusing our previous notation, we now refer to $x_{ij}(t)$ as the probability that a player with strategy i deems a player with strategy j as good at time t. When rewriting recursion (3) for this special case, we get two types of equations, depending on whether a player with strategy i meets a co-player with the same strategy or with a different one,

$$\begin{aligned} x_{ii}(t+1) &= \Big(1-w-\lambda_i(\bar{w}-w) + wr_i\big(1+\lambda_i(1-2\varepsilon)(k_i-2)\big)\Big)\cdot x_{ii}(t) \\ &\quad + w\lambda_i r_i(1-2\varepsilon)\cdot\sum_{l\neq i} k_l\cdot x_{il}(t) + \Big(w\,q_i + \lambda_i(\bar{w}-w)(\varepsilon r_i + q_i)\Big), \\ x_{ij}(t+1) &= \Big(1-w-\lambda_i(\bar{w}-w)\Big)\cdot x_{ij}(t) + wr_i\Big(1+\lambda_i(1-2\varepsilon)(k_i-1)\Big)\cdot x_{ji}(t) \\ &\quad + w\lambda_i r_i(1-2\varepsilon)\cdot\sum_{l\neq i}(k_l-1_{jl})\cdot x_{jl}(t) + \Big(w\,q_i + \lambda_i(\bar{w}-w)(\varepsilon r_i + q_i)\Big). \end{aligned} \tag{52}$$

In the above equation, the symbol 1_{jl} is an indicator function; its value is one if $j=l$ and zero otherwise. Similar to the previous section, we can rewrite the above equation as $\mathbf{x}(t+1) = \mathbf{M}\,\mathbf{x}(t)+\mathbf{v}$, where $\mathbf{M}$ is now an $s^2\times s^2$ matrix, and $\mathbf{x}(t)$ and $\mathbf{v}$ are s^2-dimensional vectors. From this recursion, we can again compute the expected probability x_{ij} to find an i-player's automaton with respect to a j-player in a good state, using Eq. (7). The payoff π_i of a player with strategy i then becomes

$$\pi_i = \sum_{j=1}^{s} \frac{k_j - 1_{ij}}{n-1} \cdot \Big(x_{ji} \cdot b - x_{ij} \cdot c\Big). \tag{53}$$

Especially when there are only a few different strategies in a large population, this algorithm can be expected to run considerably faster compared to the algorithm derived in **Section 3**.

For the limit of rare mutations and for the characterization of Nash equilibria it is useful to have explicit formulas for the special case that only $s=2$ strategies are present in the population (a resident and a deviating mutant). Suppose there are k players with strategy $\sigma_1 = (y_1, p_1, q_1, \lambda_1)$ and $n-k$ players with strategy $\sigma_2 = (y_2, p_2, q_2, \lambda_2)$. Then the matrix $\mathbf{M}$ takes the following form,

$$\mathbf{M} = \begin{pmatrix} \begin{matrix} 1-w-\lambda_1(\bar{w}-w)+ \\ wr_1\big(1+\lambda_1(1-2\varepsilon)(k-2)\big) \end{matrix} & w\lambda_1 r_1(1-2\varepsilon)(n-k) & 0 & 0 \\ 0 & 1-w-\lambda_1(\bar{w}-w) & wr_1\big(1+\lambda_1(1-2\varepsilon)(k-1)\big) & w\lambda_1 r_1(1-2\varepsilon)(n-k-1) \\ w\lambda_2 r_2(1-2\varepsilon)(k-1) & wr_2\big(1+\lambda_2(1-2\varepsilon)(n-k-1)\big) & 1-w-\lambda_2(\bar{w}-w) & 0 \\ 0 & 0 & w\lambda_2 r_2(1-2\varepsilon)k & \begin{matrix} 1-w-\lambda_2(\bar{w}-w)+ \\ wr_2\big(1+\lambda_2(1-2\varepsilon)(n-k-2)\big) \end{matrix} \end{pmatrix}$$

The vectors $\mathbf{v}$ and $\mathbf{x_0}$ are given by

$$\mathbf{v} = \Big(wq_1+\lambda_1(\bar{w}-w)(\varepsilon r_1+q_1),\ wq_1+\lambda_1(\bar{w}-w)(\varepsilon r_1+q_1),\ wq_2+\lambda_2(\bar{w}-w)(\varepsilon r_2+q_2),\ wq_2+\lambda_2(\bar{w}-w)(\varepsilon r_2+q_2)\Big)^{\intercal},$$

$$\mathbf{x_0} = (y_1, y_1, y_2, y_2)^{\intercal}.$$

The weighted average probability that an i-player considers a j-player as good again can be written as

$$\mathbf{x} = (x_{11}, x_{12}, x_{21}, x_{22}) = (\mathbf{1} - d\mathbf{M})^{-1}\big((1-d)\mathbf{x_0} + d\mathbf{v}\big).$$

The equations for the two payoffs then become

$$\begin{aligned} \pi_1 &= \left(\frac{k-1}{n-1} \cdot x_{11} + \frac{n-k}{n-1} \cdot x_{21}\right) b - \left(\frac{k-1}{n-1} \cdot x_{11} + \frac{n-k}{n-1} \cdot x_{12}\right) c, \\ \pi_2 &= \left(\frac{k}{n-1} \cdot x_{12} + \frac{n-k-1}{n-1} \cdot x_{22}\right) b - \left(\frac{k}{n-1} \cdot x_{21} + \frac{n-k-1}{n-1} \cdot x_{22}\right) c. \end{aligned} \tag{54}$$

7.2 Proofs of the equilibrium results

Proof of Lemma 1. Let us calculate the converse probability $1{-}\delta$ that a given focal pair does not interact again after it has interacted in the current round. This case occurs if either the entire population game ends immediately after the current round; or the population game continues for one more round but the focal pair does not interact in that game; or the population game continues for two more rounds without the focal pair interacting, etc. That is, we obtain

$$1{-}\delta = (1{-}d) + d(1{-}d)(1{-}w) + d^2(1{-}d)(1{-}w)^2 + \ldots = \frac{1{-}d}{1 - d(1{-}w)}. \tag{55}$$

Solving this equation for δ and using $w = 2/\big(n(n{-}1)\big)$ yields Eq. (9). □

Proof of Lemma 2. As introduced in Section 3, let $x_{ij}(t)$ denote the probability that player i considers player j to be good at time t (or equivalently, that player i would cooperate with j in round t). Assume for the moment that players apply arbitrary strategies $(y_i, p_i, q_i, \lambda_i)$. We consider the following quantity,

$$f_{ij}(T) := (1-d)\sum_{\tau=0}^{T} d^\tau \big(d \cdot x_{ij}(\tau{+}1) - x_{ij}(\tau)\big). \tag{56}$$

In this formula, we can express $x_{ij}(\tau{+}1)$ in terms of $x_{ij}(\tau)$ using Equation (1). This yields

$$\begin{aligned} f_{ij}(T) = (1-d)\sum_{\tau=0}^{T} d^\tau \cdot \Big[-\Big(1{-}d{+}dw{+}d\lambda_i(\bar{w}{-}w)\Big) \cdot x_{ij}(\tau) + dwr_i \cdot x_{ji}(\tau) \\ + d\lambda_i(1{-}2\varepsilon)r_i \cdot \sum_{l\neq i,j} wx_{jl}(\tau) + d\Big(wq_i{+}\lambda_i(\bar{w}{-}w)(\varepsilon r_i{+}q_i)\Big)\Big]. \end{aligned} \tag{57}$$

Taking the limit as T becomes large, Eq. (57) becomes

$$\begin{aligned} \lim_{T\to\infty} f_{ij}(T) = -\Big(1{-}d{+}dw{+}d\lambda_i(\bar{w}{-}w)\Big) \cdot x_{ij} + dwr_i \cdot x_{ji} \\ + d\lambda_i(1{-}2\varepsilon)r_i \cdot \sum_{l\neq i,j} wx_{jl} + d\Big(wq_i{+}\lambda_i(\bar{w}{-}w)(\varepsilon r_i{+}q_i)\Big), \end{aligned} \tag{58}$$

where $x_{ij} = (1{-}d)\sum_{\tau=0}^{\infty} d^\tau x_{ij}(\tau)$ is again the weighted average probability that player i deems player j as good, as defined in Eq. (7).

On the other hand, $f_{ij}(T)$ takes the form of a telescopic sum. So we obtain an alternative representation of the limit by calculating

$$\lim_{T\to\infty} f_{ij}(T) = \lim_{T\to\infty} (1{-}d)\Big(d^T x_{ij}(T) - x_{ij}(0)\Big) = -(1{-}d)y_i. \tag{59}$$

Because the two right hand sides of Eq. (58) and Eq. (59) need to agree, we obtain

$$\big(1-d+dw+d\lambda_i(\bar{w}-w)\big)\cdot x_{ij} - dwr_i\cdot x_{ji} - d\lambda_i(1-2\varepsilon)r_i\cdot\sum_{l\neq i,j} wx_{jl} - d\big(wq_i+\lambda_i(\bar{w}-w)(\varepsilon r_i+q_i)\big) = (1-d)y_i. \quad (60)$$

Now, due to the assumptions of the lemma we are considering a homogeneous population such that all players employ the same strategy, $(y_i, p_i, q_i, \lambda_i) = (y, p, q, \lambda)$ for all i. Due to symmetry, it follows that $x_{ij} = x_{ji} = x_{jl} =: x$ for all i, j, l. In that case, Eq. (60) simplifies to

$$\big(1-d+dw+d\lambda(\bar{w}-w)\big)x - dwrx - d\lambda(1-2\varepsilon)r(\bar{w}-w)x - d\big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\big) = (1-d)y. \quad (61)$$

Solving this equation for x yields

$$x = \frac{(1-d)y + d\big(wq + \lambda(\bar{w}-w)(\varepsilon r+q)\big)}{(1-d) + d\big(w(1-r) + \lambda(\bar{w}-w)(2\varepsilon r+1-r)\big)}. \quad (62)$$

Plugging in the definitions $w = 2/(n(n-1))$ and $\bar{w} = 2/n$, and using the expression for δ derived in Lemma 1, Eq. (62) simplifies to Eq. (10). The second part of the lemma follows by solving Eq. (10) for $x=1$ and $x=0$, respectively. □

Proof of Lemma 3. Suppose without loss of generality that the first $n-1$ individuals apply the resident strategy (y, p, q, λ), whereas player n applies some arbitrary strategy (not necessarily reactive). Then each resident enforces a relationship of the form (60), with $i \in \{1, \ldots, n-1\}$ and $j = n$,

$$\big(1-d+dw+d\lambda(\bar{w}-w)\big)x_{in} - dwrx_{ni} - d\lambda(1-2\varepsilon)rw\sum_{l\neq i,n} x_{nl} = (1-d)y + d\big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\big). \quad (63)$$

Multiplying both sides of the equation with $1/(n-1)$ and summing over all $i \in \{1, \ldots, n-1\}$ yields

$$\big(1-d+dw+d\lambda(\bar{w}-w)\big)\sum_{i=1}^{n-1}\frac{x_{in}}{n-1} - dwr\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1} - d\lambda(1-2\varepsilon)r(\bar{w}-w)\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1} = (1-d)y + d\big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\big). \quad (64)$$

By rearranging the terms in Eq. (64), we can calculate how often the first $n-1$ players cooperate on average against the mutant player,

$$\sum_{i=1}^{n-1}\frac{x_{in}}{n-1} = \frac{(1-d)y + d\big(wq+\lambda(\bar{w}-w)(\varepsilon r+q)\big)}{\big(1-d+dw+d\lambda(\bar{w}-w)\big)} + \frac{dr\big(w + \lambda(1-2\varepsilon)(\bar{w}-w)\big)}{\big(1-d+dw+d\lambda(\bar{w}-w)\big)}\sum_{i=1}^{n-1}\frac{x_{ni}}{n-1}. \quad (65)$$

Note that according to Eq. (65), there is a linear relationship between how often residents cooperate on average against the mutant, and how often the mutant cooperates on average against each resident.

According to the payoff formula (8), the mutant's payoff is given by

$$\pi_n = \sum_{i=1}^{n-1} \frac{x_{in}}{n-1} b - \sum_{i=1}^{n-1} \frac{x_{ni}}{n-1} c. \tag{66}$$

We can then plug Eq. (65) into Eq. (66). Using the definitions $w = 2/(n(n-1))$ and $\bar{w} = 2/n$, and using the formula for δ derived in Lemma 1, we obtain after some rearranging,

$$\pi_n = A_1 + A_2(r - r^*_\lambda) \sum_{i=1}^{n-1} \frac{x_{ni}}{n-1}, \tag{67}$$

where

$$\begin{aligned}
A_1 &= \frac{(1-\delta)y + \delta\big(q+(n-2)(q+r\varepsilon)\lambda\big)}{1+(n-2)\delta\lambda} \cdot b, \\
A_2 &= \frac{1+(n-2)(1-2\varepsilon)\lambda}{1+(n-2)\delta\lambda} \cdot \delta b, \\
r^*_\lambda &= \frac{1+(n-2)\delta\lambda}{1+(n-2)(1-2\varepsilon)\lambda} \cdot \frac{c}{\delta b}.
\end{aligned} \tag{68}$$

□

Proof of Theorem 1. Depending on $r = p - q$, we distinguish three cases:

(a) $r = r^*_\lambda$. In this case, Lemma 3 implies that any strategy yields the same payoff $\pi' = A_1$ against $n-1$ co-players with strategy σ. In particular, there is no profitable deviation from σ. Hence σ is a Nash equilibrium.

(b) $r < r^*_\lambda$. It follows from Lemma 3 that a strategy is a best response to σ if and only if it defects in every interaction. In particular, σ is a best response to itself if and only if a population of σ players is fully defecting. By Lemma 2, the population is fully defecting if and only if $y = p = q = 0$, or if $y = q = 0$ and either $\lambda = 0$, $n = 2$, or $\varepsilon = 0$.

(c) $r > r^*_\lambda$. For a strategy with $r > r^*_\lambda$, Lemma 3 implies that σ is a best response to itself if and only if a population of σ players is fully cooperative. While $y = p = q = 1$ is a fully cooperative strategy according to Lemma 2, this strategy does not satisfy $r = p - q > r^*_\lambda$. Hence, this case only permits a Nash equilibrium if $y = p = 1$, $q < 1 - r^*_\lambda$, and either $\lambda = 0$, $n = 2$, or $\varepsilon = 0$.

□

Proof of Theorem 2. The first two cases follow directly from Eqs. (15) and (16) by solving for $q \geq 0$. For the third case, consider a cooperative Nash equilibrium $\sigma = (1, 1, q^*_\lambda, \lambda)$ with $0 < \lambda < 1$. We calculate the partial derivative of $q^*_\lambda = 1 - r^*_\lambda$ with respect to λ,

$$\frac{\partial q^*_\lambda}{\partial \lambda} = \frac{(n-2)(1-\delta-2\varepsilon)}{\big(1+(n-2)(1-2\varepsilon)\lambda\big)^2} \cdot \frac{c}{\delta b}. \tag{69}$$

We can distinguish three cases:

(i) $\delta = 1-2\varepsilon$. In this case, we obtain $\partial q^*_\lambda/\partial\lambda = 0$ for all values of λ. In particular, if $q^*_\lambda \geq 0$ for some given $\lambda \in (0,1)$, then also $q^*_0 \geq 0$ and $q^*_1 \geq 0$.

(i) $\delta < 1-2\varepsilon$. In this case we have $\partial q^*_\lambda/\partial\lambda > 0$ for all λ. In particular, if $q^*_\lambda \geq 0$ for some given $\lambda \in (0,1)$, then $q^*_1 \geq 0$.

(i) $\delta > 1-2\varepsilon$. We conclude that $\partial q^*_\lambda/\partial\lambda < 0$ for all λ; if $q^*_\lambda \geq 0$ for $\lambda \in (0,1)$, then $q^*_0 \geq 0$.

□

Proof of Corollary 1. Suppose $\sigma = (1,1,q^*_\lambda,\lambda)$ is a cooperative Nash equilibrium for given values of n, δ and ε. We distinguish two cases:

(i) $\delta \geq 1-2\varepsilon$. It follows from the proof of Theorem 2 that also *GTFT* $= (1,1,q^*_0,0)$ is a Nash equilibrium for the given parameter values. Moreover, because the threshold value δ_0 in Eq. (17) is independent of n and ε, it follows that *GTFT* is also a cooperative Nash equilibrium for all $n' \geq n$, $\delta' \geq \delta$, and $\varepsilon' \leq \varepsilon$.

(i) $\delta < 1-2\varepsilon$. In this case it follows that *GSCO* $= (1,1,q^*_1,1)$ is also a Nash equilibrium for the given parameter values. Because the threshold δ_1 is monotonically increasing in ε and monotonically decreasing in n (for $\delta < 1-2\varepsilon$), *GSCO* is also a cooperative Nash equilibrium for all $n' \geq n$, $\delta' \geq \delta$, and $\varepsilon' \leq \varepsilon$.

□

Proof of Theorem 3. In the limit $\varepsilon \to 0$, it follows from Theorem 2 that there is a cooperative Nash equilibrium in reactive strategies if and only if

$$\delta \geq \delta_1 = \frac{c}{(n-1)b - (n-2)c}. \tag{70}$$

Now suppose to the contrary there is some arbitrary other strategy σ that is a cooperative Nash equilibrium for some $\delta < \delta_1$. In particular, players must not have an incentive to deviate by playing *ALLD*. Because players in a cooperative equilibrium need to cooperate against unknown co-players, a player who deviates to *ALLD* obtains at least the payoff of b in the first interaction he participates in. Thus, *ALLD*'s expected deviation payoff from σ satisfies

$$\pi_D \geq (1-d)b \cdot \Big(1 + (1-\bar{w})d + (1-\bar{w})^2 d^2 + \dots\Big) = \frac{(1-d)b}{1-(1-\bar{w})d}. \tag{71}$$

For σ to be a Nash equilibrium it thus needs to be the case that $\pi_D \leq b - c$, which implies

$$d \geq \frac{c}{\bar{w}b + (1-\bar{w})c}. \tag{72}$$

Using the identities $\bar{w} = 2/n$ and $d = n(n-1)\delta/\big(2 + (n-2)(n+1)\delta\big)$, this inequality simplifies to $\delta \geq \delta_1$, with δ_1 as defined in Eq. (70). □

Proof of Theorem 4. The first part of the proof of Theorem 4 is analogous to the proof of Lemma 3 and Theorem 1, and is therefore omitted here. The second part follows from the requirement that the conditional cooperation probability q needs to satisfy $q \geq 0$ for $\lambda = 0$ and $\lambda = 1$, respectively. □

Proof of Theorem 5. If implementation errors occur with a probability e, a player with strategy $\sigma = (y, p, q, \lambda)$ effectively employs the strategy $\hat{\sigma} = (\hat{y}, \hat{p}, \hat{q}, \hat{\lambda})$ where

$$\hat{y} = (1-e)y + e(1-y), \quad \hat{p} = (1-e)p + e(1-p), \quad \hat{q} = (1-e)q + e(1-q), \quad \hat{\lambda} = \lambda. \tag{73}$$

For the strategy σ to be a generic Nash equilibrium, it now needs to be the case that

$$\hat{r} := \hat{p} - \hat{q} = r^*_\lambda, \tag{74}$$

where r^*_λ is as defined in the model without implementation errors, Eq. (12). In addition, for a homogeneous σ-population to be fully cooperative in the absence of errors, we require

$$y = 1 \text{ and } p = 1. \tag{75}$$

Jointly solving Eqs. (73)–(75) for q yields

$$q = 1 - \frac{1 + (n-2)\delta\lambda}{(1-2e)\big(1 + (n-2)(1-2\varepsilon)\lambda\big)} \frac{c}{\delta b}. \tag{76}$$

This proves the first part. The second part again follows from $q \geq 0$ for both $\lambda = 0$ and $\lambda = 1$. □

Proof of Theorem 6. Again, the first part of the proof of Theorem 6 is analogous to the proof of Lemma 3 and Theorem 1; the second part follows from requiring $q \geq 0$ for $\lambda = 0$ and $\lambda = 1$, respectively. □

Proof of Theorem 7. The first part follows immediately from Lemma 3, by replacing λ by the effective probability $\lambda\nu$ to take indirect information into account. The second part again is a consequence of the requirement that $q \geq 0$ for both $\lambda = 0$ and $\lambda = 1$. □

Proof of Theorem 8. As with the previous results, the first part of the proof of Theorem 8 is analogous to the proof of Lemma 3. For example, Eq. (63) needs to be replaced by

$$\Big(1-d+wd(1-\kappa)+d(\bar{w}-w)\kappa\Big)x_{in}-dw(1-\kappa)rx_{ni}-dwr(1-2\varepsilon)\kappa\sum_{l\neq i,n}x_{nl} = (1-d)y+d\Big(w(1-\kappa)q+\kappa(\bar{w}-w)(q+\varepsilon r)\Big). \tag{77}$$

As a result, the analogous quantities to A_1, A_2, and r^*_λ in Lemma 3 are given by

$$\begin{aligned} A_1 &= \frac{(1-\delta)y+\delta\Big((1-\kappa)q+(n-2)\kappa(q+r\varepsilon)\Big)}{1+(n-3)\delta\kappa}\cdot b \\ A_2 &= \frac{1-\kappa+\kappa(n-2)(1-2\varepsilon)}{1+(n-3)\delta\kappa}\cdot \delta b \\ r^*_\kappa &= \frac{1+(n-3)\delta\kappa}{1-\kappa+\kappa(n-2)(1-2\varepsilon)}\frac{c}{\delta b}. \end{aligned} \tag{78}$$

The second part of the Theorem then follows from requiring $q := 1 - r^*_\kappa \geq 0$. □

7.3 Python code used for the evolutionary analysis

Simulation of the game dynamics for a given population composition.

```
import math, random
import numpy as np
import itertools
from itertools import islice

ctr=0
c2=0
b=5 #benefit of cooperation
c=1 #cost of cooperation
eps=float(sys.argv[1]) #error on observations of indirect interactions
delta=float(sys.argv[2]) #round continuation probability
payoffplotmat=[] #matrix to hold
payoffalld=[]
gammamat=[]
N=int(sys.argv[3]) #number of players
#defining a player as a class, with
#ID of player,
```

```
#accumulated payoff
#automata states with regard to co-players
#strategy
#IDs of connected players: everyone save for player himself
#player's average payoff over all rounds if he played at least once
class player:
    def __init__(self, playerid, payoffsum=0, states=np.zeros((N)),strat=[1,0]):
        self.playerid=playerid
        self.payoffsum = payoffsum
        self.stateslist = states
        self.strategy = strat
        self.connected =[]
        for i in range(0,N):
            if i!=self.playerid:
                self.connected.append(i)
        self.rounds=0
    def avg(self):
        if self.rounds!=0:
            return (1.*self.payoffsum)/self.rounds
        else:
            return 0


#function that advances repeated game by one round of Prisoner's Dilemma
def gamemove(player1,player2,la,ctr):
    #get automata states of focal players
    state1=player1.stateslist[player2.playerid]
    state2=player2.stateslist[player1.playerid]


    #payoffs from one round of interaction between P1 and P2,
    #according to PD payoff matrix with parameters b,c
    if (state1==0 and state2==0):
        player1.payoffsum+=0
        player2.payoffsum+=0
    elif (state1==0 and state2==1):
        player1.payoffsum+=b
        player2.payoffsum-=c
    elif (state1==1 and state2==0):
        player1.payoffsum-=c
        player2.payoffsum+=b
    else:
        player1.payoffsum+=(b-c)
        player2.payoffsum+=(b-c)



    #Both players update the state of their automaton with respect to each other,
    # according to their strategies
    if state2==1:
        if random.random()<=player1.strategy[0]:
            player1.stateslist[player2.playerid]=1
```

```
            else:
                player1.stateslist[player2.playerid]=0
        if state2==0:
            if random.random()<=player1.strategy[1]:
                player1.stateslist[player2.playerid]=1
            else:
                player1.stateslist[player2.playerid]=0

        if state1==1:
            if random.random()<=player2.strategy[0]:
                player2.stateslist[player1.playerid]=1
            else:
                player2.stateslist[player1.playerid]=0
        if state1==0:
            if random.random()<=player2.strategy[1]:
                player2.stateslist[player1.playerid]=1
            else:
                player2.stateslist[player1.playerid]=0

        #Rounds played by the two players increase by one
        player1.rounds+=1
        player2.rounds+=1

        #Players connected to focal players update their automata
        #with probability la, error eps
        #and also according to their strategies
        for i in player1.connected:
            if i!=player2.playerid:
                if random.random()<=la:
                    if state1==1:
                        if random.random()>=eps:
                            if random.random()<=playerarray[i].strategy[0]:
                                playerarray[i].stateslist[player1.playerid]=1
                            else:
                                playerarray[i].stateslist[player1.playerid]=0
                        else:
                            if random.random()<=playerarray[i].strategy[1]:
                                playerarray[i].stateslist[player1.playerid]=1
                            else:
                                playerarray[i].stateslist[player1.playerid]=0
                    if state1==0:
                        if random.random()>=eps:
                            if random.random()<=playerarray[i].strategy[1]:
                                playerarray[i].stateslist[player1.playerid]=1
                            else:
                                playerarray[i].stateslist[player1.playerid]=0
                        else:
                            if random.random()<=playerarray[i].strategy[0]:
                                playerarray[i].stateslist[player1.playerid]=1
                            else:
                                playerarray[i].stateslist[player1.playerid]=0

```

```
    for j in player2.connected:
        if j!=player1.playerid:
            if random.random()<=la:
                if state2==1:
                    if random.random()>=eps:
                        if random.random()<=playerarray[j].strategy[0]:
                            playerarray[j].stateslist[player2.playerid]=1
                        else:
                            playerarray[j].stateslist[player2.playerid]=0
                    else:
                        if random.random()<=playerarray[j].strategy[1]:
                            playerarray[j].stateslist[player2.playerid]=1
                        else:
                            playerarray[j].stateslist[player2.playerid]=0
                if state2==0:
                    if random.random()>=eps:
                        if random.random()<=playerarray[j].strategy[1]:
                            playerarray[j].stateslist[player2.playerid]=1
                        else:
                            playerarray[j].stateslist[player2.playerid]=0
                    else:
                        if random.random()<=playerarray[j].strategy[0]:
                            playerarray[j].stateslist[player2.playerid]=1
                        else:
                            playerarray[j].stateslist[player2.playerid]=0

    return

#function that starts a game round by choosing two random players to interact
def startgame(playerarray):
    while 1:
        i=random.randint(0,N-1)
        j=random.randint(0,N-1)
        if i!=j:
            break
    return playerarray[i], playerarray[j]

rounds=100000 #number of rounds
files = [open("SIM_complete_s2_{}.".format(x),'wb') for x in range(10)]

#run game for specified number of rounds,
#for 101 values of lambda from 0 to 1,
#record players' average payoffs and save them to files
for g in islice(itertools.count(),101):
    c2=0
    playerarray=[]
    #give players strategies
    for p in range(0,N):
        playerarray.append(player(p,0,np.ones(N),[1,0.01]))
    playerarray[3].strategy=[0,0]
    playerarray[3].stateslist=np.zeros(N)

```

```
    while c2<=rounds:
        payoffaverageavg=0
        ctr=0
        la=0.01*g
        for p in playerarray:
            p.stateslist=np.ones(N)
        playerarray[3].stateslist=np.zeros(N)
        currentplayers=startgame(playerarray)
        gamemove(currentplayers[0],currentplayers[1],la,ctr)
        #if random number is smaller than delta,
        #continue, otherwise, end the game and calculate average payoffs
        while 1:
            roundcont=random.random()
            if roundcont<=delta:
                currentplayers=startgame(playerarray)
                gamemove(currentplayers[0],currentplayers[1],la,ctr)
                ctr+=1
            else:
                break
        c2+=1

    #save payoffs of all players to files
    for i in range(10):
        print>>files[i],playerarray[i].avg()

for f in files:
    f.close()
```

Exact calculation of the player's payoffs for a given population composition, with k strategies.

```
import math, numpy as np, random, sys, scipy
from scipy import sparse
from scipy.sparse import linalg

n=int(sys.argv[1])
b=float(sys.argv[2])
c=float(sys.argv[3])

slist=[[[0,1,1,0.01],9],[[0,0,0,0],1]]


def rcalc(i,pvec,qvec):
    r=pvec[i]-qvec[i]
    return r

#This function calculates the index of the entries x_ji (reversed indices)
#in a vector or matrix that's indexed with k
def kprime(k,l,strat):
    return k+strat*(k%strat-k/strat)-(k%strat)+l

```

```
#This function creates the matrix M from which the vector x
#and later payoffs are calculated
def matrixcalc(kvec,lavec,yvec,pvec,qvec,delta,eps,strat):
    mat=np.zeros((strat**2,strat**2))
    rvec=[rcalc(i,pvec,qvec) for i in range(len(pvec))]
    w_ij=2./(n*(n-1))

    #populate quadratic matrix that has (strat^2) many rows and columns
    #according to algorithm in section~\ref{sec:kdifferent}
    for k in range(0,strat**2):
        i=k/strat #calculate i,j indices
        j=k%strat
        if i==j: #More than one can have same strategy
            mat[k][k]=(1.*(n-2))/n + w_ij*rvec[i] + \
            w_ij*lavec[i]*(kvec[i]-2)*rvec[i]*(1-2*eps) + \
            w_ij*(n-2)*(1-lavec[i])
            for l in range(0,strat):
                if l!=i:
                    mat[k][kprime(k,l,strat)]=w_ij*lavec[i]*\
                    kvec[l]*rvec[i]*(1 - 2*eps)
        elif i!=j:
            mat[k][kprime(k,i,strat)]=w_ij*rvec[i] + w_ij*lavec[i]*\
            (kvec[i] - 1)*rvec[i]*(1 - 2*eps) #term with i,j interchanged
            mat[k][k]=(1.*(n-2))/n+w_ij*(n-2)*(1-lavec[i])
            for l in range(0,strat):
                if (l!=j)and(l!=i):
                    mat[k][kprime(k,l,strat)]=w_ij*lavec[i]*\
                    (kvec[l])*rvec[i]*(1 - 2*eps)
                if l==j:
                    mat[k][kprime(k,l,strat)]=w_ij*lavec[i]*\
                    (kvec[l]- 1)*rvec[i]*(1 - 2*eps)

    return mat

#This function calculates the inhomogeneity in Eq.XXXX
def vcalc(kvec,lavec,fvector,pvec,qvec,delta,eps,strat):
    rvec=[rcalc(i,pvec,qvec) for i in range(len(pvec))]
    v=np.zeros(strat**2)
    w_ij=2./(n*(n-1))
    for k in range(0,strat**2): #this vector has the same entries *strat* often.
        i=k/strat
        j=k%strat
        v[k]=(lavec[i]*(n - 2)*w_ij*(qvec[i] + eps*rvec[i]) + w_ij*qvec[i])
    return v

#This function calculates payoffs given a list of strat, delta and epsilon
#by using Eq.XXXX
def payoffcalc(slist,delta,eps):
    strat=len(slist)
    kvec=[k[1] for k in slist]
    lavec=[k[0][0] for k in slist]
    yvec=[k[0][1] for k in slist]
```

```
    pvec=[k[0][2] for k in slist]
    qvec=[k[0][3] for k in slist]
    v=vcalc(kvec,lavec,fvector,pvec,qvec,delta,eps,strat)
    M2=matrixcalc(kvec,lavec,fvector,pvec,qvec,delta,eps,strat)
    x0=[]
    for k in range(0,strat**2):
        i=k/strat
        x0.append(fvector[i])
    matrixval=np.identity(strat**2)-np.dot(delta,M2)
    if strat>1:
        matrixvalsparse=sparse.csc_matrix(matrixval)
        lu = sparse.linalg.splu(matrixvalsparse)
        eye = np.eye(strat**2)
        ba = lu.solve(eye)

    else:
        ba=np.linalg.inv(matrixval)
    x=np.dot(ba,(np.dot((1-delta),x0)+np.dot(delta,v)))
    payvec=np.zeros(strat)
    for i in range(strat):
        for l in range(0,strat):
            if l!=i:
                payvec[i]+=1./(n-1)*(kvec[l]*x[strat*l+i]*b-\
                kvec[l]*x[strat*i+l]*c)
            if l==i:
                payvec[i]+=1./(n-1)*((kvec[l]-1)*x[strat*l+i]*b-\
                 (kvec[l]-1)*x[strat*i+l]*c)
    return payvec

lamatrix=[]
paymatrixcalc=[]
paymatrixalldcalc=[]
for i in range(0,101):
    la=0.01*i
    lamatrix.append(la)
    for i in range(2):
        slist[i][0][0]=la
    p=payoffcalc(slist,0.999,0.45)
    paymatrixcalc.append(p[0])
    paymatrixalldcalc.append(p[1])
```

Simulation of the strategy dynamics in an evolving population.

```
import math, numpy as np, random, sys

#Function that can calculate payoff time averages, having
#two lists (time steps list and payoff list)
def average(list1,list2):
    prd=[]
    for j in range(0,len(list1)):
```

```
        prd.append(list1[j]*list2[j])
    average=sum(prd)/sum(list1)
    return average

#Function that calculates p-q
def rcalc(s):
    r=s[2]-s[3]
    return r

#Function that calculates payoffs in a population with 2 strategies
#of type (y,p,q,la),
#where k players play strategy 1
def payoffcalc(k,s1,s2,delta,eps):
    r1=rcalc(s1)
    r2=rcalc(s2)
    la1=s1[0]
    la2=s2[0]
    w_ij=2./(n*(n-1))
    m11=(1.*(n-2))/n + w_ij*r1 + \
    w_ij*la1*(k-2)*r1*(1-2*eps) + w_ij*(n-2)*(1-la1)
    m12=w_ij*la1*(n - k)*r1*(1 - 2*eps)
    m13=0
    m14=0
    m21=0
    m22=(1.*(n-2))/n+w_ij*(n-2)*(1-la1)
    m23=w_ij*r1 + w_ij*la1*(k - 1)*r1*(1 - 2*eps)
    m24=w_ij*la1*(n - k - 1)*r1*(1 - 2*eps)
    m31=w_ij*la2*(k - 1)*r2*(1 - 2*eps)
    m32=w_ij*r2 + w_ij*la2*(n - k - 1)*r2*(1 - 2*eps)
    m33=w_ij*(n-2)*(1-la2)+(1.*(n-2))/n
    m34=0
    m41=0
    m42=0
    m43=w_ij*la2*k*r2*(1 - 2*eps)
    m44=(1.*(n - 2))/n + w_ij*(n - 2)*(1 - la2) +\
    w_ij*r2 + w_ij*la2*(n - k - 2)*r2*(1 - 2*eps)
    M=np.array([[m11,m12,m13,m14],[m21,m22,m23,m24],\
    [m31,m32,m33,m34],[m41,m42,m43,m44]])
    M2=M #.T
    for i in range(4):
        if i<2:
            v[i]=(la1*(n - 2)*w_ij*(s1[3] + eps*r1) + w_ij*s1[3])
        else:
            v[i]=(la2*(n - 2)*w_ij*(s2[3] + eps*r2) + w_ij*s2[3])

    x0=np.array([s1[1],s1[1],s2[1],s2[1]])
    #print x0
    #print M2
    #delta=0.85
    matrixval=np.identity(4)-np.dot(delta,M2)
    #print matrixval
    x=np.dot(np.linalg.inv(matrixval),(np.dot((1-delta),x0)+np.dot(delta,v)))
```

```
    #x=np.dot(np.linalg.inv(matrixval),v)
    #print x
    payoff1 =1./(n-1)*(((k-1)*x[0]+(n-k)*x[2])*b-((k-1)*x[0]+(n-k)*x[1])*c)
    payoff2=1./(n-1)*((k*x[1]+(n-k-1)*x[3])*b-(k*x[2]+(n-k-1)*x[3])*c)
    return payoff1, payoff2

#Function that calculates exponential term
#in formula for the fixation probability
def db(i,s1,s2,delta,eps):
    p=payoffcalc(i,s1,s2,delta,eps)
    db=math.exp(-s*(p[0]-p[1]))
    return db

#Function that calculates fixation probability
#of a mutant strategy in resident strategy
def fixprob(s1,s2,delta,eps):
    l=0.
    a=1.
    for i in range(1,n):
        a*=db(i,s1,s2,delta,eps)
        #print a
        l+=a
    xf=1./(1+l)
    return xf

#running evolutionary simulation for *rounds*, with parameters
#delta (global round continuation probability),
#eps (error on indirect information),
#b (benefit of cooperation) and c (cost of cooperation),
#s (strength of selection)
def main(delta, eps, n, b, c, s, rounds)
        s1=np.zeros(4)
        s2=np.zeros(4)
        ctr=0
        e=epsstart
        stratcounter=1
        savearray=[]
        switch=0
        for i in range(rounds):
        s1[0]=random.randint(0,1)
        s1[1]=random.random()
        s1[2]=random.random()
        s1[3]=random.random()
        y1=fixprob(s1,s2,d,eps)
        z=random.random()
        if z<=y1:
            #save datapoints in array that will be reused only
            #if strategy fixes in population
            savearray.append([[s2[0],s2[1],s2[2],s2[3]],\
            payoffcalc(0,s1,s2,d,eps)[1],stratcounter])
            stratcounter=1
            s2[0]=s1[0]
```

```
            s2[1]=s1[1]
            s2[2]=s1[2]
            s2[3]=s1[3]
            switch+=1
        else:
            #strategy is not switched, so increase its counter
            stratcounter+=1
        i+=1
        savearray.append([[s2[0],s2[1],s2[2],s2[3]],\
        payoffcalc(0,s1,s2,d,eps)[1],stratcounter])
        #save trajectories to files
        payofflist=[row[1] for row in savearray]
        lalist=[row[0][0] for row in savearray]
        tlist=[row[2] for row in savearray]
        #average payoff and receptivity for parameter combination
        payoffavg=average(tlist,payofflist)
        laavg=average(tlist,lalist)
```

8 Supplementary References

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

[2] Molander, P. The optimal level of generosity in a selfish, uncertain environment. *Journal of Conflict Resolution* **29**, 611–618 (1985).

[3] Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).

[4] Kraines, D. P. & Kraines, V. Y. Pavlov and the prisoner's dilemma. *Theory and Decision* **26**, 47–79 (1989).

[5] Imhof, L. A., Fudenberg, D. & Nowak, M. A. Tit-for-tat or win-stay, lose-shift? *Journal of Theoretical Biology* **247**, 574–580 (2007).

[6] van Segbroeck, S., Pacheco, J. M., Lenaerts, T. & Santos, F. C. Emergence of fairness in repeated group interactions. *Physical Review Letters* **108**, 158104 (2012).

[7] Fischer, I. *et al.* Fusing enacted and expected mimicry generates a winning strategy that promotes the evolution of cooperation. *PNAS* **110**, 10229–10233 (2013).

[8] Pinheiro, F. L., Vasconcelos, V. V., Santos, F. C. & Pacheco, J. M. Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Comput Biol* **10**, e1003945 (2014).

[9] Rapoport, A., Seale, D. A. & Colman, A. M. Is Tit-for-Tat the answer? On the conclusions drawn from axelrod's tournaments. *PLoS One* **10**, e0134128 (2015).

[10] Lee, C., Harper, M. & Fryer, D. The art of war: Beyond memory-one strategies in population games. *PLoS One* **10**, e0120625 (2015).

[11] Dong, Y., Li, C., Tao, Y. & Zhang, B. Evolution of conformity in social dilemmas. *PLoS One* **10**, e0137435 (2015).

[12] Boyd, R. & Lorberbaum, J. No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. *Nature* **327**, 58–59 (1987).

[13] Farrell, J. & Ware, R. Evolutionary stability in the repeated prisoner's dilemma. *Theoretical Population Biology* **36**, 161–166 (1989).

[14] Boyd, R. Mistakes allow evolutionary stability in the repeated Prisoner's Dilemma game. *Journal of Theoretical Biology* **136**, 47–56 (1989).

[15] Fudenberg, D. & Maskin, E. Evolution and cooperation in noisy repeated games. *American Economic Review* **80**, 274–279 (1990).

[16] Bendor, J. & Swistak, P. Types of evolutionary stability and the problem of cooperation. *Proceedings of the National Academy of Sciences USA* **92**, 3596–3600 (1995).

[17] van Veelen, M., García, J., Rand, D. G. & Nowak, M. A. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA* **109**, 9929–9934 (2012).

[18] García, J. & van Veelen, M. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory* **161**, 161–189 (2016).

[19] Hauert, C. & Schuster, H. G. Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. *Proceedings of the Royal Society B* **264**, 513–519 (1997).

[20] Baek, S. K., Jeong, H. C., Hilbe, C. & Nowak, M. A. Comparing reactive and memory-one strategies of direct reciprocity. *Scientific Reports* **6**, 25676 (2016).

[21] Stewart, A. J. & Plotkin, J. B. Small groups and long memories promote cooperation. *Scientific Reports* **6**, 26889 (2016).

[22] Hilbe, C., Martinez-Vaquero, L. A., Chatterjee, K. & Nowak, M. A. Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences USA* **114**, 4715–4720 (2017).

[23] García, J. & Traulsen, A. The structure of mutations and the evolution of cooperation. *PLoS One* **7**, e35287 (2012).

[24] Chatterjee, K., Zufferey, D. & Nowak, M. A. Evolutionary game dynamics in populations with different learners. *Journal of Theoretical Biology* **301**, 161–173 (2012).

[25] Stewart, A. J. & Plotkin, J. B. The evolvability of cooperation under local and non-local mutations. *Games* **6**, 231–250 (2015).

[26] García, J. & van Veelen, M. No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI* **5**, 102 (2018).

[27] Hilbe, C., Chatterjee, K. & Nowak, M. A. Partners and rivals in direct reciprocity. *Nature Human Behaviour* **2**, 469–477 (2018).

[28] Press, W. H. & Dyson, F. D. Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *PNAS* **109**, 10409–10413 (2012).

[29] Stewart, A. J. & Plotkin, J. B. Extortion and cooperation in the prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* **109**, 10134–10135 (2012).

[30] Hilbe, C., Nowak, M. A. & Sigmund, K. The evolution of extortion in iterated prisoner's dilemma games. *Proceedings of the National Academy of Sciences USA* **110**, 6913–6918 (2013).

[31] Stewart, A. J. & Plotkin, J. B. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proceedings of the National Academy of Sciences USA* **110**, 15348–15353 (2013).

[32] Hilbe, C., Nowak, M. A. & Traulsen, A. Adaptive dynamics of exortion and compliance. *PLoS One* **8**, e77886 (2013).

[33] Szolnoki, A. & Perc, M. Evolution of extortion in structured populations. *Physical Review E* **89**, 022804 (2014).

[34] Szolnoki, A. & Perc, M. Defection and extortion as unexpected catalysts of unconditional cooperation in structured populations. *Scientific Reports* **4**, 5496 (2014).

[35] Hilbe, C., Wu, B., Traulsen, A. & Nowak, M. A. Cooperation and control in multiplayer social dilemmas.

Proceedings of the National Academy of Sciences USA **111**, 16425–16430 (2014).
[36] Hao, D., Rong, Z. & Zhou, T. Extortion under uncertainty: Zero-determinant strategies in noisy games. *Physical Review E* **91**, 052803 (2015).
[37] Liu, J., Li, Y., Xu, C. & Hui, P. M. Evolutionary behavior of generalized zero-determinant strategies in iterated prisoner's dilemma. *Physica A* **430**, 81–92 (2015).
[38] Akin, E. The iterated prisoner's dilemma: Good strategies and their dynamics. In Assani, I. (ed.) *Ergodic Theory, Advances in Dynamics*, 77–107 (de Gruyter, Berlin, 2016).
[39] McAvoy, A. & Hauert, C. Autocratic strategies for iterated games with arbitrary action spaces. *Proceedings of the National Academy of Sciences* **113**, 3573–3578 (2016).
[40] McAvoy, A. & Hauert, C. Autocratic strategies for alternating games. *Theoretical Population Biology* **113**, 13–22 (2016).
[41] Xu, X., Rong, Z., Wu, Z.-X., Zhou, T. & Chi, K. T. Extortion provides alternative routes to the evolution of cooperation in structured populations. *Physical Review E* **95**, 052302 (2017).
[42] Ichinose, G. & Masuda, N. Zero-determinant strategies in finitely repeated games. *Journal of Theoretical Biology* **438**, 61–77 (2018).
[43] Boerlijst, M. C., Nowak, M. A. & Sigmund, K. Equal pay for all prisoners. *American Mathematical Monthly* **104**, 303–307 (1997).
[44] Reiter, J. G., Hilbe, C., Rand, D. G., Chatterjee, K. & Nowak, M. A. Crosstalk in concurrent repeated games impedes direct reciprocity and requires stronger levels of forgiveness. *Nature Communications* **9**, 555 (2018).
[45] Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
[46] Alexander, R. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
[47] Kandori, M. Social norms and community enforcement. *Review of Economic Studies* **59**, 63–80 (1992).
[48] Okuno-Fujiwara, M. & Postlewaite, A. Social norms and random matching games. *Games and Economic Behavior* **9**, 79–109 (1995).
[49] Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577 (1998).
[50] Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Proceedings of the Royal Society B* **268**, 745–753 (2001).
[51] Panchanathan, K. & Boyd, R. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology* **224**, 115–126 (2003).
[52] Brandt, H. & Sigmund, K. The logic of reprobation: Assessment and action rules for indirect reciprocation. *Journal of Theoretical Biology* **231**, 475–486 (2004).
[53] Ohtsuki, H. & Iwasa, Y. How should we define goodness? – Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* **231**, 107–20 (2004).
[54] Ohtsuki, H. & Iwasa, Y. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology* **239**, 435–444 (2006).
[55] Pacheco, J. M., Santos, F. C. & Chalub, F. A. C. C. Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Computational Biology* **2**, 1634–1638 (2006).
[56] Uchida, S. & Sigmund, K. The competition of assessment rules for indirect reciprocity. *Journal of Theoretical Biology* **263**, 13–19 (2009).
[57] Martinez-Vaquero, L. A. & Cuesta, J. A. Evolutionary stability and resistance to cheating in an indirect reciprocity model based on reputation. *Physical Review E* **87**, 052810 (2013).
[58] Uchida, S., Yamamoto, H., Okada, I. & Sasaki, T. A theoretical approach to norm ecosystems: Two adaptive

architectures of indirect reciprocity show different paths to the evolution of cooperation. *Frontiers in Physics* **6**, 14 (2018).

[59] Santos, F. P., Santos, F. C. & Pacheco, J. M. Social norm complexity and past reputations in the evolution of cooperation. *Nature* **555**, 242–245 (2018).

[60] Nowak, M. A. & Sigmund, K. Evolution of indirect reciprocity. *Nature* **437**, 1291–1298 (2005).

[61] Sigmund, K. Moral assessment in indirect reciprocity. *Journal of Theoretical Biology* **299**, 25–30 (2012).

[62] Okada, I. A review of theoretical studies on indirect reciprocity. *Games* **11**, 27 (2020).

[63] Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *Journal of Theoretical Biology* **194**, 561–574 (1998).

[64] Sasaki, T., Okada, I. & Nakai, Y. The evolution of conditional moral assessment in indirect reciprocity. *Scientific Reports* **7**, 41870 (2017).

[65] Ohtsuki, H. Reactive strategies in indirect reciprocity. *Journal of Theoretical Biology* **227**, 299–314 (2004).

[66] Berger, U. Learning to cooperate via indirect reciprocity. *Games and Economic Behavior* **72**, 30–37 (2011).

[67] Berger, U. & Grüne, A. On the stability of cooperation under indirect reciprocity with first-order information. *Games and Economic Behavior* **98**, 19–33 (2016).

[68] Clark, D., Fudenberg, D. & Wolitzky, A. Individualistic records robustly support social cooperation. *Working paper* (2019).

[69] Hilbe, C., Schmid, L., Tkadlec, J., Chatterjee, K. & Nowak, M. A. Indirect reciprocity with private, noisy, and incomplete information. *Proceedings of the National Academy of Sciences USA* **115**, 12241–12246 (2018).

[70] Raub, W. & Weesie, J. Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology* **96**, 626–654 (1990).

[71] Pollock, G. & Dugatkin, L. A. Reciprocity and the emergence of reputation. *Journal of Theoretical Biology* **159**, 25–37 (1992).

[72] Roberts, G. Evolution of direct and indirect reciprocity. *Proceedings of the Royal Society B* **275**, 173–179 (2007).

[73] Nakamaru, M. & Kawata, M. Evolution of rumours that discriminate lying defectors. *Evolutionary Ecology Research* **6**, 261–283 (2004).

[74] Seki, M. & Nakamaru, M. A model for gossip-mediated evolution of altruism with various types of false information by speakers and assessment by listeners. *Journal of Theoretical Biology* **407**, 90–105 (2016).

[75] Tanabe, S., Suzuki, H. & Masuda, N. Indirect reciprocity with trinary reputations. *Journal of Theoretical Biology* **317**, 338–347 (2013).

[76] Nakamura, M. & Masuda, N. Indirect reciprocity under incomplete observation. *PLoS Computational Biology* **7**, e1002113 (2011).

[77] Brandt, H. & Sigmund, K. The good, the bad and the discriminator - errors in direct and indirect reciprocity. *Journal of Theoretical Biology* **239**, 183–194 (2006).

[78] Uchida, S. Effect of private information on indirect reciprocity. *Physical Review E* **82**, 036111 (2010).

[79] Santos, F. P., Santos, F. C. & Pacheco, J. M. Social norms of cooperation in small-scale societies. *PLoS Computational Biology* **12**, e1004709 (2016).

[80] Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).

[81] Hilbe, C., Traulsen, A. & Sigmund, K. Partners or rivals? Strategies for the iterated prisoner's dilemma. *Games and Economic Behavior* **92**, 41–52 (2015).

[82] Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Physical Review*

E **74**, 011909 (2006).

[83] Imhof, L. A. & Nowak, M. A. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B* **277**, 463–468 (2010).

[84] Fudenberg, D. & Imhof, L. A. Imitation processes with small mutations. *Journal of Economic Theory* **131**, 251–262 (2006).

[85] Wu, B., Gokhale, C. S., Wang, L. & Traulsen, A. How small are small mutation rates? *Journal of Mathematical Biology* **64**, 803–827 (2012).

[86] Traulsen, A. & Hauert, C. Stochastic evolutionary game dynamics. In Schuster, H. G. (ed.) *Reviews of Nonlinear Dynamics and Complexity*, 25–61 (Wiley-VCH, Weinheim, 2009).

[87] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).

[88] Wu, B., García, J., Hauert, C. & Traulsen, A. Extrapolating weak selection in evolutionary games. *PLoS Computational Biology* **9**, e1003381 (2013).

[89] Wang, Z. & Durrett, R. Extrapolating weak selection in evolutionary games. *Journal of Mathematical Biology* **78**, 135–154 (2019).

[90] Traulsen, A., Hauert, C., De Silva, H., Nowak, M. A. & Sigmund, K. Exploration dynamics in evolutionary games. *Proceedings of the National Academy of Sciences USA* **106**, 709–712 (2009).

[91] Hilbe, C. Local replicator dynamics: A simple link between deterministic and stochastic models of evolutionary game theory. *Bulletin of Mathematical Biology* **73**, 2068–2087 (2011).

[92] Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A. & Sigmund, K. Via freedom to coercion: the emergence of costly punishment. *Science* **316**, 1905–1907 (2007).

[93] Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466**, 861–863 (2010).

[94] Hilbe, C. & Traulsen, A. Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. *Scientific Reports* **2**, 458 (2012).

[95] Nowak, M. A., Sigmund, K. & El-Sedy, E. Automata, repeated games and noise. *Journal of Mathematical Biology* **33**, 703–722 (1995).

[96] Posch, M. Win-stay, lose-shift strategies for repeated games - memory length, aspiration levels and noise. *Journal of Theoretical Biology* **198**, 183–195 (1999).