



City Research Online

City St George's, University of London

Citation: Wei, M., Wang, X., Liu, L., Zheng, H., Jiang, Y., Hao, Y., Zheng, Z., Fu, F. & Tang, S. (2025). Indirect reciprocity in the public goods game with collective reputations. *Journal of The Royal Society Interface*, 22(225), 20240827. doi: 10.1098/rsif.2024.0827

This is the published version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/35599/>

Link to published version: <https://doi.org/10.1098/rsif.2024.0827>

Copyright and Reuse: Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).



Research



Cite this article: Wei M, Wang X, Liu L, Zheng H, Jiang Y, Hao Y, Zheng Z, Fu F, Tang S. 2025 Indirect reciprocity in the public goods game with collective reputations. *J. R. Soc. Interface* **22**: 20240827.

<https://doi.org/10.1098/rsif.2024.0827>

Received: 20 November 2024

Accepted: 21 January 2025

Subject Category:

Life Sciences—Mathematics interface

Subject Areas:

evolution, biomathematics

Keywords:

collective reputation, game theory, indirect reciprocity, social norm

Authors for correspondence:

Xin Wang

e-mail: wangxin_1993@buaa.edu.cn

Shaoting Tang

e-mail: tangshaoting@buaa.edu.cn

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7669668>.

Indirect reciprocity in the public goods game with collective reputations

Ming Wei^{1,2}, Xin Wang^{2,3,4}, Longzhao Liu^{2,3,4}, Hongwei Zheng⁵, Yishen Jiang^{1,2}, Yajing Hao^{1,2}, Zhiming Zheng^{2,3,4}, Feng Fu⁶ and Shaoting Tang^{2,3,4}

¹School of Mathematical Sciences, ²LMIB, NLSDE, Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing, and ³School of Artificial Intelligence, Beihang University, Beijing, People's Republic of China

⁴Zhongguancun Laboratory, Beijing, People's Republic of China

⁵Beijing Academy of Blockchain and Edge Computing, Beijing, People's Republic of China

⁶Department of Mathematics, Dartmouth College, Hanover, NH, USA

MW, 0009-0000-6626-5480; FF, 0000-0001-8252-1990

Indirect reciprocity unveils how social cooperation is founded upon moral systems. Within the frame of dyadic games based on individual reputations, the 'leading-eight' strategies distinguish themselves in promoting and sustaining cooperation. However, in real-world societies, there are widespread interactions at the group level, where individuals need to make a singular action choice when facing multiple individuals with different reputations. Here, through introducing the assessment of collective reputations, we develop a framework that embeds group-level reputation structure into public goods games to study the evolution of group-level indirect reciprocity. We show that changing the criteria of group assessment destabilizes the reputation dynamics of leading-eight strategies. In a particular range of social assessment criteria, all leading-eight strategies can break the social dilemma in public goods games and sustain cooperation. Specifically, there exists an optimal, moderately set assessment criterion that is most conducive to promoting cooperation. Moreover, in the evolution of assessment criteria, the preference of the leading-eight strategies for social strictness is inversely correlated with the payoff level. Our work reveals the impact of social strictness on prosocial behaviour, highlighting the importance of group-level interactions in the analysis of evolutionary games and complex social dynamics.

1. Introduction

Cooperation among individuals serves as a crucial foundation for the development of human societies. Although altruistic behaviours may not be the most advantageous choices for individuals in the short term, they foster community solidarity and collective long-term benefits. To maintain cooperation, many societies have developed reputation-based moral systems to regulate social behaviour, commonly referred to as social norms [1–4]. Many social norms have been recognized as essential for promoting societal advancement [5–7]. Within the context of these norms, individuals enhance their reputations by engaging in prosocial behaviours, enabling them to receive rewards [7–10] or avoid punishments [11–14] in subsequent social interactions. This mechanism, whereby indirect information such as reputation influences individual behaviour, is known as indirect reciprocity, which has been proven to promote cooperation in social dilemmas [15]. Recent reviews also highlight how moral preferences and language-based interactions drive behavioural changes in social systems, affecting cooperation [16,17]. Within the framework of evolutionary game theory [6,18,19],

		Reputations of donor and recipient(s)				
		GG	GB	BG	BB	
Action of donor	C	G	L1/L3/L4/L7 L2/L5/L6/L8	G	L1/L2/L3/L5 L4/L6/L7/L8	Assessment Rule
	D	B	G	B	L3/L4/L5/L6 L1/L2/L7/L8	
		C	D	C	L1/L2 L3-L8	Action Rule

Figure 1. We denote the leading-eight strategies as L1 through L8. Each strategy includes an assessment rule and an action rule corresponding to a full set of binary values across the boxes shown. The assessment rule regards an action as good (G) or bad (B) according to the context information and the action itself, and the action rule guides the donor to cooperate (C) or defect (D) according to the reputations of the game participants. Each column represents a reputation combination for the donor and the recipient. For instance, BG indicates bad donors versus good recipients. Here, good assessments and cooperative actions are marked in blue, and bad assessments and defective actions are marked in red. The leading-eight strategies behave identically in certain scenarios (boxes with a single letter), especially when facing interactions involving good recipients. In contrast, disagreements emerge when they deal with bad passive parties, leading to different phenomena in evolutionary dynamics.

researchers have explored how indirect reciprocity affects the evolution of cooperation, identifying eight effective strategies for maintaining cooperation. These strategies are referred to as the leading-eight strategies (figure 1), including the well-known Consistent Standing (L2), Simple Standing (L3), Stern Judging (L6), Staying (L7) and Judging (L8) [20,21]. Common features of these strategies, such as cooperation with good individuals and opposition to the betrayal of such individuals, reflect findings from empirical studies on social norms within real communities across different cultural contexts [4,5,22,23]. Therefore, exploring the performance of the leading-eight strategies within the context of evolutionary game theory holds significant importance for understanding the development of cooperation in real-world societies.

So far, researchers have drawn various conclusions regarding the leading-eight strategies. These strategies have been shown to effectively maintain cooperation in environments characterized by public assessments [20,21]. When private assessments are employed within a population, the stability of the leading-eight strategies is weakened to varying degrees [24,25]. Nonetheless, implementing appropriate quantitative assessments in place of binary reputations can mitigate the impact of private information on the stability of the leading-eight strategies [26]. Moreover, some studies have focused on the scope of information sharing in indirect reciprocity, exploring the evolution of different social norms [27] and the impact of stereotypes on indirect reciprocity [28]. In addition, a recent study suggests that assessing individuals based on several behaviours allows for a more accurate capture of behavioural patterns, thereby facilitating agreement among individuals [29]. These results make significant contributions to the literature on indirect reciprocity. However, they have mostly focused on pairwise interactions. Their models adopt the precondition that both the donor and the recipient consist of only one individual, respectively. In real societies, group interactions are prevalent [30,31]. As members of a group, individuals' behaviours and cognition are influenced by group-level indirect information, which is also subjected to certain social norms. Therefore, it is natural to pose the question: How do social norms influence human behaviour in the context of group interactions?

In fact, cooperative behaviour at the group level has been studied through various approaches [32–35]. Public goods games are commonly employed as one of the standard paradigms to explore cooperation in multi-player interactions [36,37]. Players are usually assumed to take binary actions, typically cooperation or defection. Their actions may be affected by the environment [38–41], the size of the game [42], the reward system [43], the spatial and network structures [44–47], and the recently widely considered higher-order interactions [48]. To date, some studies have explored the dynamic mechanisms of indirect reciprocity based on group-level interactions. In public goods games [49] or three-player donation games [50], the strict strategy that cooperates only when all other players have good credit can promote indirect reciprocity. Initiated with one round of the public goods game, pairwise interactions can stabilize cooperation without the second-order free-rider problem [51]. These studies have reached qualitatively different conclusions from those focused on dyadic games. However, as an abstract model of social norms that emerge in real societies, the impact of the leading-eight strategies on the evolution of cooperation in group interactions remains unknown.

To this end, we propose a framework for collective reputation assessment based on the leading-eight strategies to study the evolutionary process of indirect reciprocity in public goods games. Individuals naturally appraise a group according to the current reputations of its members. Within this reputation structure, one intriguing factor is the proportion of good reputations within the group. The threshold of it, which we refer to as the group assessment criterion, indicates the minimum fraction of the members who must have good reputations in order for the group to be considered good. This framework helps us explore the dynamics of cooperation under various social norms over a spectrum of social strictness. We show how this simple mechanism qualitatively changes the evolutionary outcomes of indirect reciprocity. As the strictness of group assessment increases, the leading-eight strategies experience varying degrees of reputation instability when competing with unconditional strategies. In strict populations, none of them can effectively distinguish between cooperators and defectors. The introduction of group assessment also breaks the boundary between the phases of unconditional cooperation and defection. With moderate assessment criteria, players of the leading-eight strategies can screen out unjust defectors and maintain cooperation in social dilemmas. Moreover, by exploring the

evolution of assessment criteria, we show that the preference of the leading-eight strategies for social strictness is inversely correlated with the payoff level, and the attitude of the strategy towards cooperation between good donors and bad recipient groups determines the sensitivity of this correlation.

2. Model

Consider a well-mixed population of size N with binary reputations. To simulate the dynamics of cooperation, members of this population repeatedly go through public goods games. In each round, k players are randomly selected from the population to form a group G . Each player in G decides whether to contribute a cost c to the public wealth pool, which will eventually be distributed to members of the group after being multiplied by a synergy factor R [52]. Without loss of generality, we set $c = 1$. The act of contributing is interpreted as cooperation while the opposite is considered defection. The game keeps no memory of the players' behaviour; that is, the public pool is divided equally among all the players regardless of whether they cooperate. The number of cooperators in a round is denoted as n_C ; thus, we derive the payoffs for the cooperators and the defectors in a game with k players as $\pi_C(k, n_C) = Rcn_C/k - c$ and $\pi_D(k, n_C) = Rcn_C/k$, respectively. Individual reputations are binary variables assigned to each ordered pair of players. To record the reputations within the population, we naturally employ the concept of image matrices [25,53,54]. For such a matrix $M(t)$, an entry $M_{ij}(t)$ equals 1 if player i considers j to be good at time t ; otherwise, the value is 0. Although a public goods game is a situation with no asymmetric status, each player engaged can be regarded as a donor, while the remaining players together act as the passive party, i.e. the recipient group. We assume that players assess a group according to the reputations of its members. To assess the reputation of the recipient group, we introduce a novel reputation structure termed collective reputation. Specifically, player i regards group G' as good if, from i 's perspective, the proportion of good reputations inside G' is at least λ , namely

$$r_{G'}^i = \begin{cases} 1, & \sum_{j \in G'} M_{ij}(t) \geq \lambda |G'| \\ 0, & \sum_{j \in G'} M_{ij}(t) < \lambda |G'| \end{cases} \quad (2.1)$$

where λ is referred to as the group assessment criterion, which describes the strictness of the player. Note that, although all players receive the same payoff in public goods games, separating the actor and the recipient group allows our model to recover the reputation dynamics of the two-player game with $|G'| = 1$ and $\lambda \in (0, 1]$ (electronic supplementary material), and this setup does not qualitatively affect the outcomes. To interpret other players' actions, each individual is equipped with a strategy $S = \{r(r_a, b, r_p), b(r_a, r_p)\}$, consisting of assessment rules (the first term) and action rules (the second term). r_a and r_p represent the reputations of the active party (the donor) and the passive party (the recipients), respectively, while b represents the behaviour observed by the strategy holder. Each variable and function introduced above is binary, with values of 0 (for bad/defection) or 1 (for good/cooperation). Before taking action, a donor i inside group G evaluates the reputation of the donor's corresponding recipient $G(i) = G \setminus \{i\}$, denoted as $r_{G(i)}^i$. To highlight the impact of group structure on information transmission, we assume that players participating in the game always learn each other's actions. Outside the group, each individual observes the game with probability $q > 0$. Upon observation, the action of each player might be mistakenly interpreted by an observer with probability $\epsilon \geq 0$. In this case, a cooperative action C may be seen as a D , or a D as a C , analogously. Player h assesses i based on not only i 's previous reputation and action but also the collective reputation of $G(i)$. Then, the entry $M_{hi}(t)$ is adjusted to $M_{hi}(t + 1)$. After the game, the reputation of donor i is updated across the population. As each player involved in the group should be considered as the donor, we use a synchronous updating scheme, applying this process to each member in G equivalently before moving on to the next round of the game (figure 2).

3. Results

3.1. Analysis of collective reputation dynamics

We first focus on how group assessment influences reputation dynamics with different social norms. For simplicity, we consider scenarios in which individuals are unable to change their strategies. In each case, the population is evenly distributed among three strategies: one of the leading-eight (Li), unconditional cooperation (ALLC) and unconditional defection (ALLD). The leading-eight players in a population share the same value of group assessment criterion λ , thus forming a unified level of strictness in the society. Roughly speaking, as λ gradually increases from 0 to 1, the society transitions from relaxed to moderate and then to strict. Notably, the perspectives of ALLC and ALLD players are of little importance because, regardless of the recipient's reputation, these players always stick to their deterministic actions. Therefore, in each scenario, the key to reputation dynamics lies in analysing the attitudes of the leading-eight towards the three strategies. Without loss of generality, we assume that ALLC players always consider people to be good and that ALLD players always consider people to be bad. The leading-eight strategies have been shown to behave differently under such population compositions with dyadic interactions [24]. Regarding group interactions, however, more interesting results emerge concerning changes in the group assessment criterion.

In a relaxed society (with small λ), all strategies except ALLD cooperate actively, making it easy for leading-eight players to distinguish these free riders. However, with an increase in strictness, the stability of reputations is affected to varying degrees (figure 3). First, the reputations of the leading-eight populations fluctuate with the social strictness (figure 3a). As the society transitions from relaxed to moderately strict (moderate λ), the average proportions of good reputations in the self-assessment of the

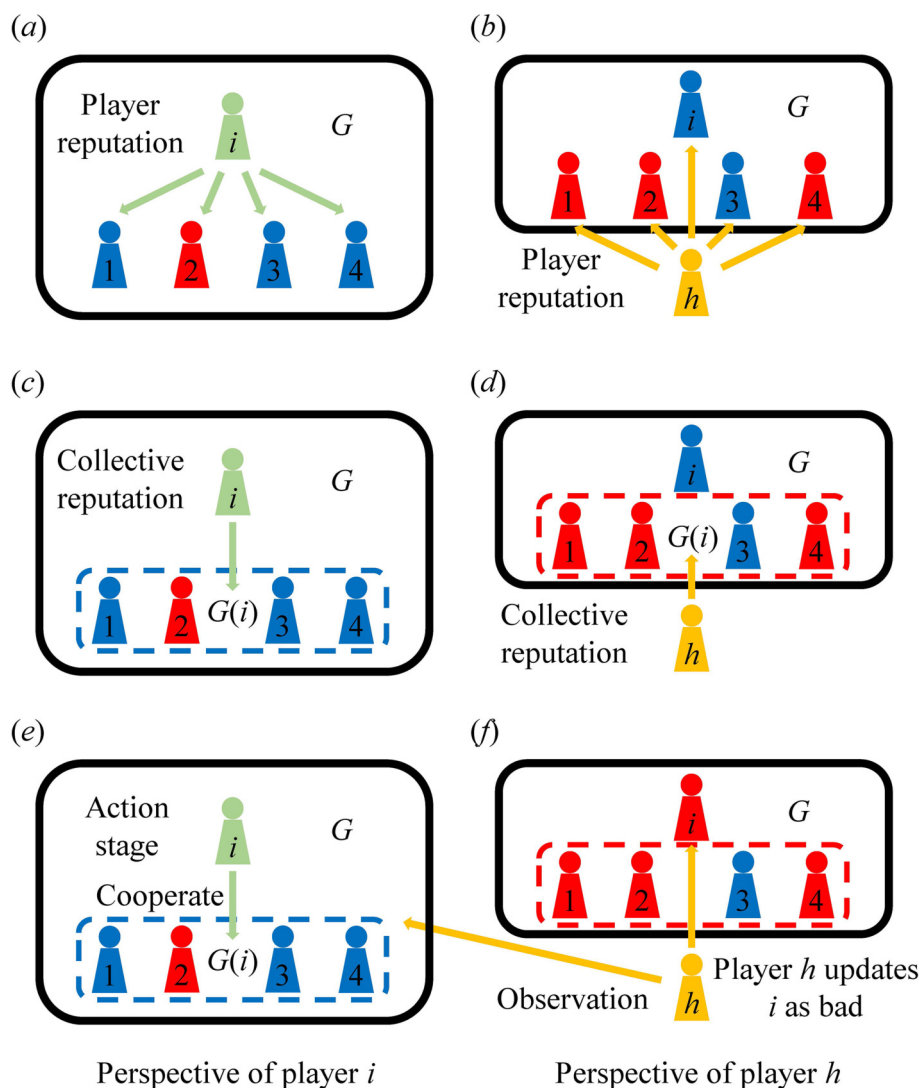


Figure 2. Schematic illustration of the modelling framework. In public goods games, players are engaged with equal status. However, from the perspective of indirect reciprocity, we can split the game into several asymmetric games in which each player acts as the donor in turn. Consider a game G of size 5 where player i acts as the donor. Player h is an observer outside the game group. (a–d) To implement the strategies, i and h need to evaluate the reputations of the recipient group $G(i) = G \setminus \{i\}$ (i.e. the set of players in G excluding i), according to their current opinions towards its members. Here, player i considers three of the four members to be good, leading to a good view of $G(i)$. Player h , on the other hand, holds different opinions towards players 1 and 4 and therefore dislikes $G(i)$. This process requires a threshold for the group to be considered good, which we refer to as the group assessment criterion λ . Disagreements caused by information non-transparency and strategy differences may further diverge and result in a series of nontrivial phenomena. In this case, i and h have contrary views of $G(i)$ due to their disagreements regarding the group's members. (e) Players take action once they are selected to participate in a game. Their choices are observed by individuals outside the group and influenced by a certain extent of private or noisy information. In this stage, i cooperates with $G(i)$, and i 's action is learned by h as indirect information. (f) Finally, h updates the reputation of i according to h 's own knowledge of the context of the game and the action. As $G(i)$ appears untrustworthy in h 's opinion, cooperation with it spoils i 's standing.

leading-eight strategies decrease. Nonetheless, this is not always a monotonic trend. For scenarios of L1–L6, the lowest value of this proportion is not associated with the strictest criterion but corresponds to a moderate value of λ . As conditional cooperators, leading-eight players support each other in a relaxed society. The recipient groups are typically perceived as good here and cooperation becomes the dominant behaviour. On the other hand, when the criterion is quite strict, most of the leading-eight players tend to not only defect but also be empathetic for defections since good recipient groups barely exist in this case. Therefore, moderate societies with neither too relaxed nor too strict group assessment criteria surprisingly result in the greatest disruptions to social consensus. Without clear guidance for individuals to be relaxed or strict, their private information plays a more crucial role in reputation assessment. Therefore, when perception errors occur, moderate criteria will further amplify the divergence of opinions. Unlike other strategies, however, L7 and L8 seem to be extremely vulnerable to strict criteria. While L7 players exhibit a continuous decrease in self-evaluation as the strictness increases, L8 players rarely assign good reputations to any individual when $\lambda \geq 0.4$.

Next, in terms of unconditional strategies, reputation dynamics display simple but still interesting patterns. When assessing ALLC players, the leading-eight exhibit two different patterns (figure 3b). L2, L5, L6 and L8 do not appreciate cooperating with bad recipients. These strategies lose trust in ALLC and hesitate to cooperate in moderate and strict societies. In contrast, L1, L3, L4 and L7 appreciate cooperation. Good donors can maintain their standing even when they cooperate with a bad passive party. Thus, ALLC players keep their standing no matter how strict the society is, suggesting that these leading-eight strategies might

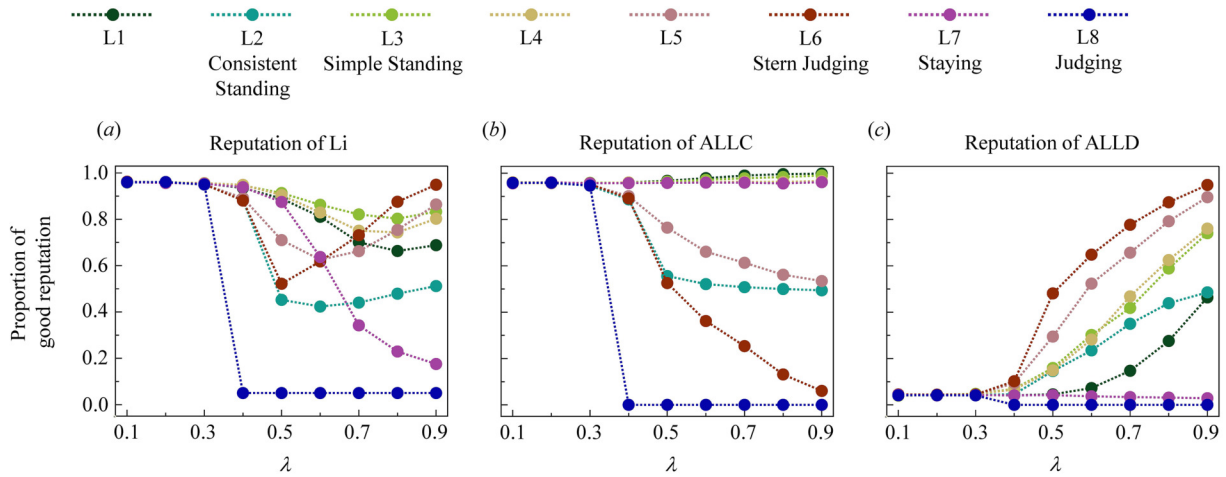


Figure 3. Changes in group assessment criterion cause divergence in the reputation dynamics. Consider the setting where each population consists of three types of strategies in equal proportions: one of the leading-eight (Li), ALLC and ALLD. This figure illustrates the proportion of good reputations among players of (a) Li, (b) ALLC and (c) ALLD as a function of the assessment criterion λ . Each point in the figure shows the average result of 50 repeated experiments, with each experiment encompassing 2×10^5 iterated steps in populations of size $N = 60$. The size of the game k is fixed to 10. Group interactions are observed with probability $q = 0.9$. Perception errors occur with probability $\epsilon = 0.05$.

be able to compete with ALLC over a wider range of social strictness. Figure 3c shows that for most of the leading-eight strategies, the reputations of ALLD accumulate as the group assessment criterion increases. Populations of L3, L4, L5 and L6 most evidently follow this pattern. These four strategies assign good reputations to more than 70% of ALLD players when λ equals 0.9. In these scenarios, a bad donor can re-establish a good reputation by defecting against bad recipients. On the other hand, L1 and L2 almost randomly assign reputations to ALLD players in strict societies while L7 and L8 nearly always despise these defectors. From the perspective of these four strategies, bad donors cannot regain a good reputation through defection. Therefore, we may expect them to be more competitive against ALLD in social dilemmas.

In summary, the leading-eight strategies share a common sense of assessment in relaxed societies. They distinguish ALLC along with their respective players from ALLD. However, as the criteria become stricter, the stability of reputations diverges. These results illustrate the importance of social tolerance in the development of reputation systems. Note that here we consider the results under the condition of a certain error rate. Although this situation is complex, under the assumption of rare errors, we can theoretically analyse the recovery process of the leading-eight population from a single error, which fits well with the corresponding simulation results (electronic supplementary material, Recovery analysis from a single error).

3.2. Evolutionary dynamics

Participants in social interactions might change their norms in the real world. Therefore, in this section, we study the evolutionary dynamics of cooperation when strategies are not fixed. To align with the previous stage, we focus on a simplified scenario. Players choose among only three strategies, including one of the leading-eight (Li), ALLC and ALLD. The process of strategy alternation follows classic simple imitation dynamics [24,55–57]. For each step, one player i is randomly selected from the population. This player may adopt some new strategy with probability μ (known as the mutation rate) or randomly pick another player j as a role model with the remaining probability $1 - \mu$. In the latter case, player i may imitate the strategy of player j with a probability $P(\bar{\pi}_i, \bar{\pi}_j) = \{1 + \exp[-s(\bar{\pi}_j - \bar{\pi}_i)]\}^{-1}$, where s denotes the strength of selection [58]. These two types of strategy updating schemes are used to construct the evolutionary process based on mutation and selection. Notably, evolutionary dynamics occur on a larger timescale compared with reputation dynamics [24,55,56]. For each possible population composition, the reputation dynamics are iterated over 2×10^5 time steps. The point is to determine how often a strategy would, on average, cooperate in group interactions under different circumstances. Based on these results of cooperation frequencies, we can further calculate the expected payoff of this strategy under each population composition. Therefore, instead of employing simulations to go through the evolutionary process, we use a Markov state transition matrix to calculate the selection–mutation equilibrium (see the electronic supplementary material for these calculations). Hereinafter, we present the results of strategy evolution with relatively rare mutations, denoted as $\mu \rightarrow 0$ [59,60]. In the context of evolutionary dynamics, this limitation implies that the evolution time ranges of two mutations do not overlap. That is, before the evolutionary process resulting from one mutation converges to a stable state (a homogeneous state), the next mutation will not occur. With such a limitation, the evolutionary system will only transition among a few homogeneous states, where the population consists of only one strategy. In the electronic supplementary material, we further discuss the results of high mutation rates.

We first focus on the cooperation rate of the population through the course of evolution. Here, we refer to the synergy factor R as a function of group size, namely $R(k) = \alpha k$, with payoff parameter α [48]. Considering cases involving substantial teamwork, such as the co-authorship of scientific publications, it has been proven that the size of a group significantly affects the function and outcomes of its activities [48,61,62]. Our motivation here is to describe the evolutionary process of group interaction, with group size as a concern. We investigate the eight scenarios over a λ – α panel, exploring the coupling effect of group assessment

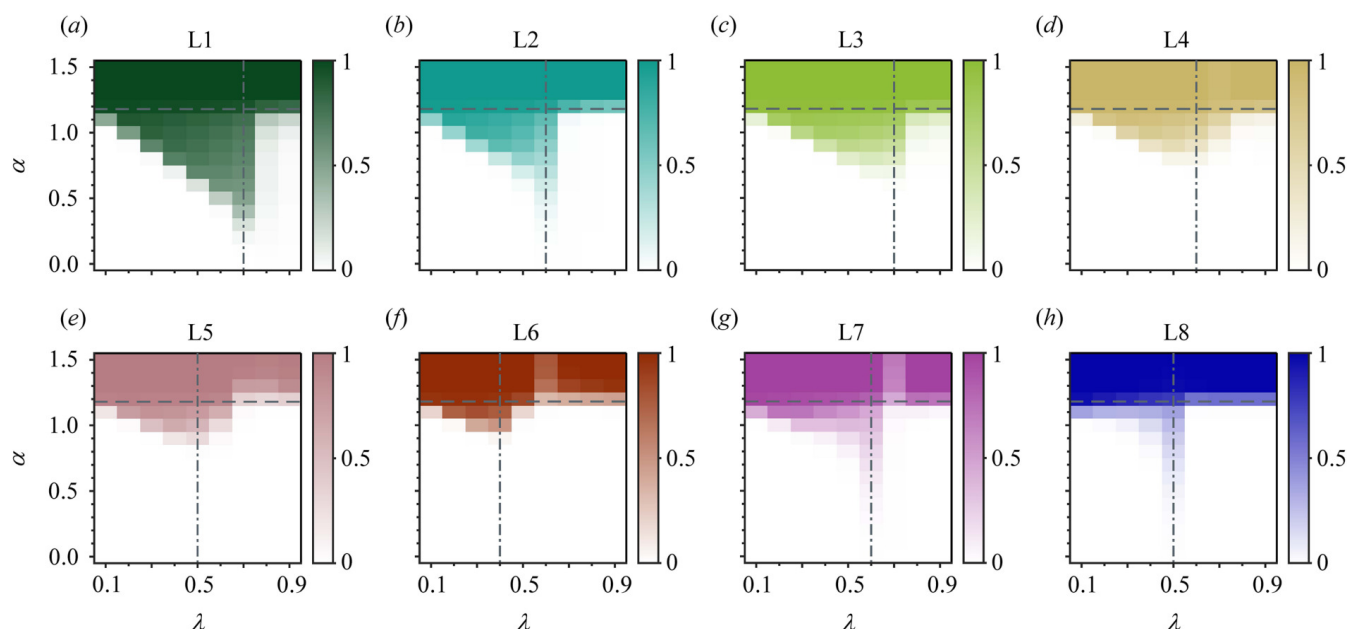


Figure 4. An optimal, moderately set group assessment criterion is most conducive to promoting cooperation in social dilemmas. Consider populations consisting of players with one of the leading-eight strategies (Li), ALLC and ALLD. (a–h) The cooperation rate of the population over the spectrum of payoff parameters α and group assessment criteria λ . The value $\alpha_c = 1.18$ depicted with the horizontal dashed line in each panel represents the threshold for unconditional cooperation to evolve. As the society turns from relaxed to moderate, cooperation emerges under conditions of $\alpha < \alpha_c$, forming a gradient boundary above the defection area (shown in white). This trend ceases around the critical value λ_c (vertical dashed-dotted line), which indicates where the leading-eight strategies are most capable of sustaining cooperation in social dilemmas. Parameters: population size $N = 60$, group size $k = 10$, selection strength $s = 1$, perception error rate $\epsilon = 0.05$, observation probability $q = 0.9$ and mutation rate $\mu \rightarrow 0$.

criterion and payoff level. The structural nature of the public goods game gives rise to a critical value of α , denoted as α_c , that separates the defection and cooperation phases (appendix A). Typically, cooperation is predominant in cases with a sufficiently high payoff ($\alpha > \alpha_c$). When $\alpha < \alpha_c$, defection becomes a more rational choice for individuals, leading to a social dilemma. However, the intervention of the group assessment criterion breaks the boundary between these two phases (figure 4). In all eight scenarios, cooperation can be maintained in social dilemmas within a range of social strictness. As λ increases from 0.1 to a moderate value, cooperation gradually emerges in lower payoff environments, forming a downward-sloping boundary above the pure defection area. This moderate value, denoted as λ_c , indicates the criterion that is most favourable for sustaining cooperation in social dilemmas for a certain leading-eight strategy (appendix B). It plays the role of a threshold, separating the range of strictness in which the leading-eight population tends towards cooperation ($\lambda < \lambda_c$) or defection ($\lambda > \lambda_c$). L1, L2, L7 and L8 with strictness λ_c are able to maintain cooperation even when α turns below 0.5 (figure 4a,b,g,h). Their harsh judgement of defections between bad parties provides them with excellent competitiveness against ALLD (figure 5a,b,g,h), resulting in this intriguing phenomenon. When social strictness exceeds λ_c , cooperation almost immediately retreats to the region above α_c , and the impact of the assessment criterion becomes less pronounced. Intuitively, the outstanding role of λ_c in promoting cooperation may be caused by the fact that moderate assessment criteria provide the leading-eight players with better abilities to identify free riders compared with relaxed or strict criteria. In relaxed societies, ALLD players can go unpunished and achieve high payoffs in the game, which gives them a competitive advantage. On the other hand, when λ is too large, justified defections of leading-eight players are more likely to be mistakenly punished. From this perspective, a moderate level of strictness can effectively restrict the exploitation gains of free riders while maintaining cooperation among leading-eight players, thereby enhancing the competitiveness of the leading-eight strategy in social dilemmas (electronic supplementary material, Cooperation in social dilemma).

The results of how often each strategy is played in the selection–mutation equilibrium and the fixation probabilities between different strategies can help us gain deeper insight (figure 5). Evolution outcomes of ALLC and ALLD are solely influenced by the payoff level, with each positioned on opposite sides of the critical value α_c . On the other hand, the leading-eight strategies are subject to the coupling effect of both the game structure and the collective reputation structure (figure 5a–h). The λ – α panels are divided into several sections based on the frequency of each strategy dominating the population in the selection–mutation equilibrium. In a low-payoff ($\alpha < \alpha_c$) environment, ALLC is not favoured, and only the leading-eight strategies can co-exist with ALLD and sustain cooperation. Nevertheless, there is a notable difference in their capabilities in this aspect. L3–L6 clearly struggle to compete with ALLD. These four strategies consider bad donors who defect against bad recipients as good, leading to their inability to distinguish and punish ALLD players. However, they still exhibit a slight evolutionary trend as λ increases from relaxed values to λ_c (figure 5c–f). On the other hand, L1, L2, L7 and L8 contend with ALLD in a larger region in different ways (figure 5a,b,g,h). L1 and L2 support cooperation between bad parties. They can enhance their own payoff through cooperation while identifying unjustified defectors. Therefore, they distinctly evolve in low-payoff scenarios both below and above λ_c . In contrast, L7 and L8 mainly co-exist with ALLD using strict criteria, and they hardly maintain any cooperation therein. The fixation probabilities between the leading-eight and each unconditional strategy are shown in figure 5i–p. The payoff parameter employed here is $\alpha = 0.9$

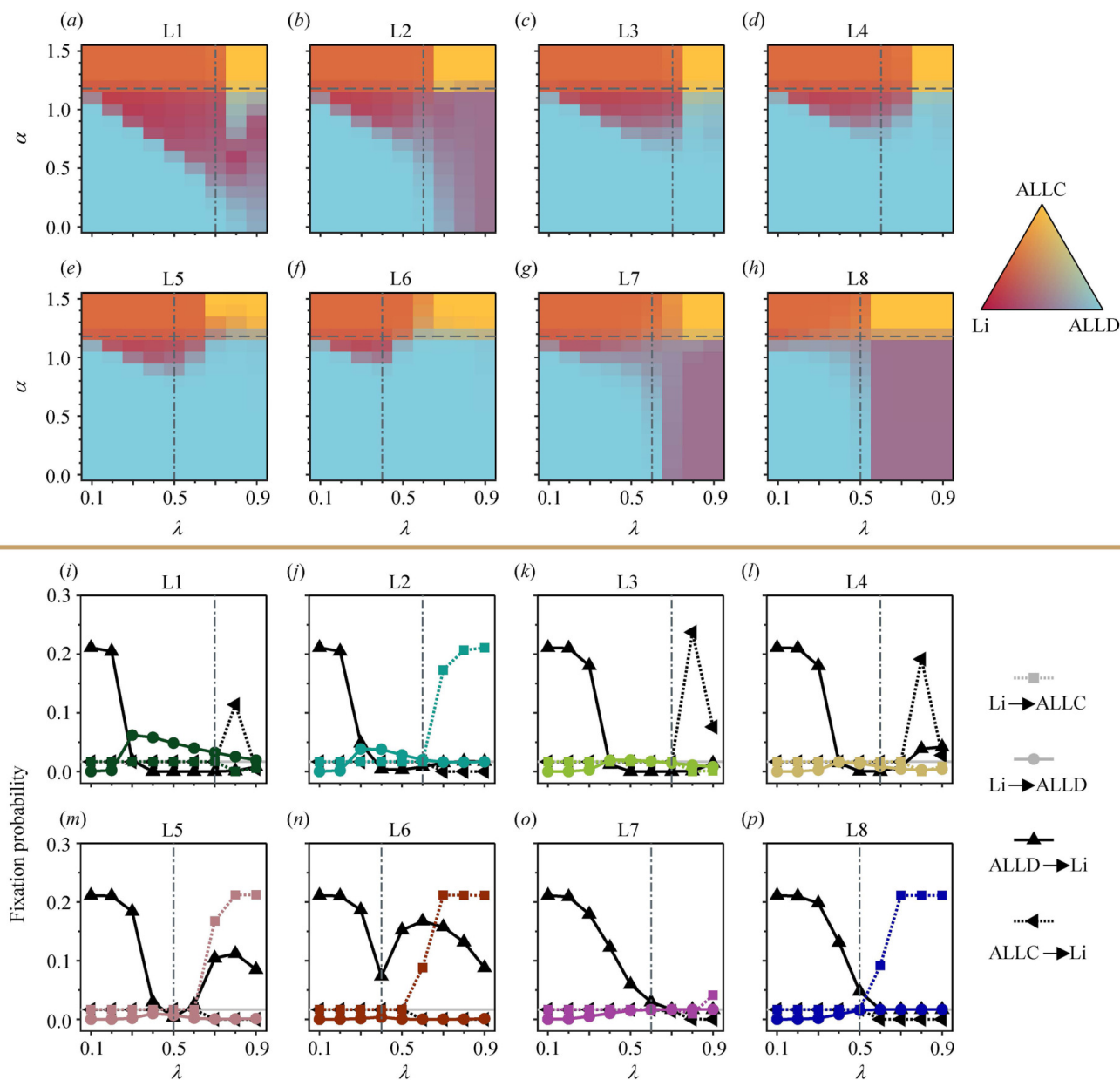


Figure 5. The evolution of the leading-eight strategies is subject to the coupling effect of the payoff parameter and group assessment. As in figure 4, the horizontal dashed line in (a–h) represents $\alpha_c = 1.18$, and the value of λ_c corresponds to the vertical dashed-dotted line in each panel. (a–h) The combined average frequency of three strategies (the leading-eight, ALLC and ALLD) in the selection–mutation equilibrium. Each coloured brick represents the weighted average of all evolutionary stable states. The triangle legend depicts the corresponding proportion of each strategy in the population. To further explain the results when $\alpha < \alpha_c$, (i–p) shows the fixation probabilities between each leading-eight strategy and both of the unconditional strategies, given $\alpha = 0.9$. The legend items signify that mutations of the strategy on the left side of the arrow fixate into the resident strategy on the right side. The solid horizontal grey line in each panel marks the value of $1/N$, which corresponds to the fixation probability of a neutral mutant. With the limitation of rare mutations, the evolutionary process transitions among several homogeneous states. Each time a mutation occurs, it may fixate into the population and replace the resident strategy or go extinct. The baseline parameters are the same as in figure 4.

to fulfil the condition of a typical social dilemma. In each panel, we highlight the fixation probability of a neutral mutant, $1/N$, with a horizontal grey line. When the criterion is relaxed, ALLD mutations are able to fixate into the resident populations of all the leading-eight strategies. However, as the strictness becomes moderate, the advantage of ALLD significantly weakens. In some scenarios, the fixation probability of ALLD against the leading-eight is even lower than $1/N$. This explains why several leading-eight strategies, although they cannot substantially invade ALLD (only L1 and L2 have fixed probabilities against ALLD exceeding $1/N$), can still co-exist with ALLD in social dilemmas.

In high-payoff ($\alpha > \alpha_c$) environments, when the criterion is relaxed or moderate (basically $\lambda \leq \lambda_c$), all of the leading-eight strategies break even with ALLC at a similar ratio. As society becomes strict, ALLC dominates the population. Notably, L2, L5, L6 and L8 oppose cooperation between good donors and bad recipients, which leads to an interesting phenomenon. In low-payoff areas where $\alpha < \alpha_c$, they can invade ALLC using strict criteria (figure 5j,m,n,p). While in high-payoff areas, ALLC dominates them more easily (at a smaller value of λ) than the other four strategies (figure 5b,e,f,h). The differing performances on either side of α_c reflect the qualitative changes in L2, L5, L6 and L8, suggesting their high sensitivity to changes in payoff levels.

Additionally, in the main text, we adopt the baseline parameter settings of $k = 10$, $\epsilon = 0.05$, $q = 0.9$ and $s = 1$. In the electronic supplementary material, we provide results under different parameter values to demonstrate the robustness of the qualitative conclusions.

3.3. Evolution of the assessment criterion

So far, we have discussed the performance of the leading-eight in competition with unconditional strategies, considering different payoff parameters and social strictness. In real-world societies, however, people may not practice ALLC or ALLD as their strategies. Instead, most individuals may be willing to cooperate under some, but not all, circumstances. This raises the following question: How should we choose our assessment criterion under different levels of social payoff? In this section, we consider populations that include only one of the leading-eight strategies that are equipped with three different values of the group assessment criterion: $\lambda_1 = 0.1$, $\lambda_2 = 0.4$ and $\lambda_3 = 0.7$. Players are allowed to alter their criteria during the evolutionary process, following the same imitation dynamics as in the previous section. The point is to study how often each criterion is used in the final state of the evolutionary process.

In low-payoff environments, the leading-eight players tend to choose a relatively strict assessment criterion. As the payoff level increases, individuals are more inclined to adopt moderate or relaxed criteria (figure 6). More specifically, the pattern of criterion evolution is determined by how the leading-eight strategy views good players who cooperate with bad recipients. For strategies L1, L3, L4 and L7, the change in payoff levels has little impact on the preferences of the population for different degrees of strictness (figure 6a,c,d,g). The three assessment criteria are essentially neutral in these four scenarios, as they all demonstrate a similarly high tendency towards cooperation (figure 6i,k,l,o). These four strategies encourage cooperation between good donors and bad recipients, ensuring that even with the strict criterion λ_3 , the reputation stability of good individuals is not compromised by cooperative behaviour, thereby facilitating the evolution of cooperation. On the other hand, for strategies L2, L5, L6 and L8, the preference for different criteria is clearly more sensitive to changes in α (figure 6b,e,f,h). The strict criterion λ_3 dominates the other two criteria and deterministically takes over the population when α is not sufficiently large. In the scenario of L6, λ_1 and λ_2 cannot fixate in the population until $\alpha > \alpha_c$. For these four strategies, cooperation between good donors and bad recipients is discouraged. Moreover, at low levels of payoff, the presence of social dilemmas puts cooperators at a disadvantage. Therefore, within the range of $\alpha < \alpha_c$, strict invaders can easily fixate into a relaxed criterion by exploiting the kindness of the resident players (figure 6j,m,n,p). The corresponding fixation probabilities of λ_3 to λ_1 and λ_2 exceed 40% for L2 and L5, and 60% for L6 and L8 when α approaches 0. As α exceeds α_c , we observe the dominance of relaxed assessment criteria over the strict one in three of the four strategies, namely L2, L5 and L8 (figure 6j,m,p). The above results of the evolution of assessment criteria are also reflected in the cooperation rate, demonstrating differences in the two types of leading-eight strategies (electronic supplementary material, figure S16). Regardless of the changes in α , L1, L3, L4 and L7 consistently exhibit high cooperation rates. On the other hand, the cooperation rates of L2, L5, L6 and L8 exhibited high sensitivities to changes in the payoff environment.

The previous experiments were based on the assumption that mutations are rare, where the population eventually stabilizes in a situation in which each individual adopts the same strategy with the same value of the group assessment criterion. However, past research has attached great importance to understanding the effects of mutations. In the electronic supplementary material (Mutation rate in assessment criterion evolution), we further explore the corresponding evolutionary results at high mutation rates. In these scenarios, populations may evolve into more complicated states in which the coexistence of several different criteria is possible. We show that the dominant correlations between different values of λ are mitigated as mutations become significant. For strategies L2, L5, L6 and L8, relaxed populations are able to persist under low payoff levels. For strategies L1, L3, L4 and L7, the three assessment criteria almost completely become neutral. To further study the impact of collective reputation in more complex scenarios, we also discuss situations in which assessment criteria and strategies coevolve within a population. We include leading-eight strategies with three different assessment criteria, along with ALLC and ALLD, to explore the dynamics of cooperation evolution in social dilemmas. In this case, the cooperation rate of the population in all eight scenarios monotonically increases with the payoff level (electronic supplementary material, Co-evolution of strategy and assessment criterion).

4. Discussion

People interact based on moral systems in social activities. Cooperation arises and evolves through reputation assessment and subsequent actions. During this process, individuals employ different strategies that rely on indirect information to make decisions, embodying the social norms. Among all strategies, the leading-eight strategies have been demonstrated to maintain cooperation with synchronized information [20,21]. These strategies share several common properties that match the moral consensus in human societies, including kindness, retaliation, apology and forgiveness. In this study, we propose an indirect reciprocity framework based on collective reputation and investigate the performance of the leading-eight strategies in the public goods game. We find that a relaxed society favours the stability of the reputation mechanism but does not support the evolution of cooperation under low payoff levels. Instead, under moderate assessment criteria, all of the leading-eight strategies can substantially sustain cooperation in social dilemmas. Furthermore, as the expected payoffs increase, the population tends to adopt more relaxed assessment criteria, which further shapes the dynamics of cooperation. Our model echoes the concept of cultural tightness–looseness in social psychology [63,64], where societies range from ‘loose’ (relaxed) to ‘tight’ (strict). Harrington and Gelfand’s analysis of the 50 United States [65] suggests that regions facing ecological or social challenges tend to become stricter, consistent with the negative correlation between λ and α . They also find that tight states exhibit higher levels of social stability yet face higher incarceration rates, greater discrimination and increased inequality. This pattern aligns with our results, highlighting that an overly relaxed

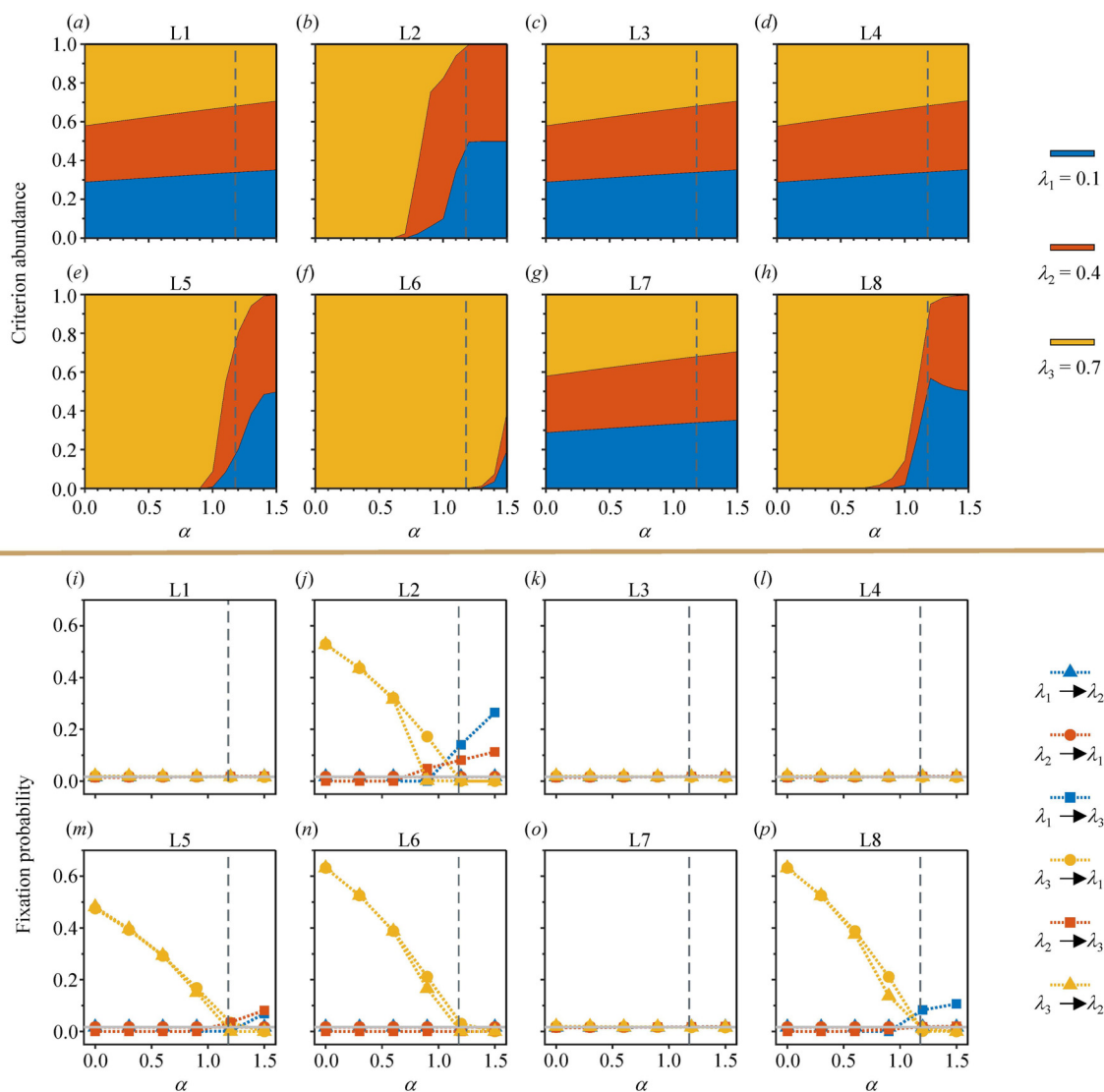


Figure 6. The average abundance and fixation probability in the evolution of assessment criteria. The vertical dashed line in each panel represents the critical payoff level $\alpha_c = 1.18$. (a–h) Average abundance for three values of λ within a spectrum of payoff parameter α . The strict criterion has better performance at low payoff levels. As the payoff increases, relaxed criteria can help players achieve higher returns through more cooperation, thereby eliminating the advantage of the strict criterion. Panels (i–p) illustrate the fixation probabilities among three different assessment criteria. The horizontal grey line in each panel indicates the neutral fixation probability $1/N$. The legend items indicate the situations in which the mutants on the left fixate into residents on the right. Parameters are the same as in figure 4. The values of λ include 0.1, 0.4 and 0.7.

or overly strict assessment criterion can be suboptimal for sustaining cooperation, whereas a balanced level of social strictness proves more conducive to long-term prosocial behaviour. Our framework disentangles the roles of game structure and reputation structure in the context of indirect reciprocity involving multiple players. The results show that collective reputation assessment can mitigate social dilemmas in public goods games. This finding offers us new insights into the evolutionary dynamics of indirect reciprocity in social dilemmas, emphasizing the importance of interpreting mesoscale social behaviour beyond the individual perspective.

The previous works on indirect reciprocity have assumed that individuals interact in a dyadic manner. In such a case, bad reputations always lead to defections. This makes it difficult for most of the leading-eight strategies to distinguish between disagreements caused by errors and true defectors, thereby failing to gain an advantage in competition with ALLD in social dilemmas. According to Hilbe *et al.* [24], only L1, L2 and L7 can maintain cooperation. However, in our reputation structure, considering group-level information, the bad reputation of a single recipient does not necessarily result in the donor's defection, which provides the leading-eight players a buffer zone to distinguish between free riders and justified defectors. By adopting moderate assessment criteria, the leading-eight strategies can effectively restrict the exploitation benefits of ALLD players while maintaining internal cooperation. This characteristic of collective reputation enables all of the leading-eight strategies to resist the invasion of ALLD to some extent in social dilemmas (electronic supplementary material, Cooperation in social dilemma).

Typically, for indirect reciprocity, the transmission of information can be divided into two types: public assessment and private assessment. In public assessment, game information is broadcasted by an observer to other individuals, leading to a unified assessment of a particular individual by the population [20,21]. Whereas in private assessment, each individual independently observes the game process and forms their own assessments. Previous research has shown that with the existence of noisy information, private assessment weakens the stability of indirect reciprocity, leading to a decrease in cooperation rates [24]. Nevertheless, with

diverse channels for obtaining firsthand information at present, we believe that private assessment may be a more appropriate way to describe the realistic social environment. Therefore, in the results of this study, we adopted private assessment. Additionally, we have also conducted a robustness analysis of error recovery under private assessment to support this choice (electronic supplementary material, Recovery analysis from a single error).

The motivation of this work is to provide a theoretical framework for exploring the dynamics of indirect reciprocity in group interactions. Therefore, we adopt the assumption of homogeneity for both groups and individuals, which enhances the flexibility and portability of our model. To avoid direct reciprocity resulting from repeated interactions among the same individuals, we consider dynamic game groups that change over time. However, it should be noted that in different scenarios, collective reputations and group interactions can be formulated in a myriad of ways. For example, leaders within a group may have more influence than other members, thereby exerting a greater impact on the collective reputation of the group. Also, group membership may be fixed or dynamic. Within the former case, images of certain groups may be more difficult to change due to stereotypes [66] or biases [67]. By incorporating relevant heterogeneity mechanisms into the framework and adjusting parameters accordingly, these new assumptions may yield interesting conclusions with different qualitative characteristics.

Note that, in evolutionary dynamics, we employ Markov chains to obtain the stable state by calculating the eigenvectors of the state transition matrix. Given that the timescale of evolutionary dynamics is significantly longer than that of reputation dynamics, this method enables us to achieve results more quickly and accurately compared to directly simulating the evolutionary process. Although we have adopted this method for well-mixed populations, we believe that the concept of group assessment can also be examined in populations with spatial structures such as lattices and networks [48]. Moreover, allowing λ to evolve continuously over $[0, 1]$ on top of the current assessment criterion framework could yield more nuanced outcomes. However, this would require substantial modifications to both the model structure and state space, which we regard as an important direction for future work.

Additionally, while our work is based on agent-based experiments, scenarios of infinite populations with replicator dynamics might be promising for analytically determining the nature of the critical criteria [50,68,69]. Modifications to the game process, such as punishment [11–13], reward [43] and quantitative assessment [26], may also alter the assessment and action rules, leading to interesting phenomena in group-level interactions.

Ethics. This work did not require ethical approval from a human subject or animal welfare committee.

Data accessibility. All simulations and numerical calculations were performed with MATLAB R2021b. The custom code that supports the findings of this study is included as electronic supplementary material and is also available at GitHub [70].

Electronic supplementary material is available online [71].

Declaration of AI use. We have not used AI-assisted technologies in creating this article.

Authors' contributions. M.W.: conceptualization, formal analysis, methodology, visualization, writing—original draft, writing—review and editing; X.W.: conceptualization, formal analysis, funding acquisition, writing—review and editing; L.L.: formal analysis; H.Z.: formal analysis, funding acquisition; Y.J.: formal analysis; Y.H.: formal analysis; Z.Z.: funding acquisition, supervision; F.F.: formal analysis, writing—review and editing; S.T.: formal analysis, funding acquisition, supervision.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. This work is supported by National Science and Technology Major Project (2022ZD0116800), Program of National Natural Science Foundation of China (62141605, 12425114, 12201026, 12301305), the Fundamental Research Funds for the Central Universities, and Beijing Natural Science Foundation (Z230001).

Appendix A. Critical value of α_c

In public goods games, when α is small, social dilemmas arise and defections become the dominant behavior; whereas when α is large, cooperation becomes the dominant behavior. The critical value that divides these two phases is denoted as $\alpha_c = (N - 1)/(N - k)$, indicating the condition under which the expected payoffs of ALLC and ALLD in the population are equal (see electronic supplementary material for detailed calculations). In the main text, under the conditions of $N = 60$ and $k = 10$, the value of α_c equals 1.18. In the electronic supplementary material, we further analysed the calculation of α_c with dynamic group sizes, based on a more complex synergy factor.

Appendix B. Definition of the critical value of assessment criterion

To accurately describe the promotion of cooperation in cases with moderate strictness, we define the critical value of the assessment criterion, λ_c , as the value of λ that most effectively promotes the evolution of cooperation in social dilemmas. For a certain leading-eight strategy, consider the cooperation rate as a function of the payoff parameter and assessment criteria, $\theta = \theta(\alpha, \lambda)$. Given a threshold θ^* of the cooperation rate, we define α^* as the minimum possible value of α corresponding to a cooperation rate exceeding θ^* , that is, $\alpha^* = \operatorname{argmin}_{\alpha} \{\theta(\alpha, \lambda) > \theta^*, \exists \lambda\}$. Then, λ_c is defined as the λ value that corresponds to the highest cooperation rate in the case of α^* , namely,

$$\lambda_c = \lambda(\alpha^*) = \operatorname{argmax}_{\lambda} \{\theta(\alpha^*, \lambda)\}. \quad (\text{B } 1)$$

The values of λ_c in the results of the main text and electronic supplementary material correspond to $\theta^* = 0.1$. Note that with θ^* in the range of $(0.01, 0.27)$, the results of λ_c are generally robust. In the electronic supplementary material, we provide the results for $\theta^* = 0.2$ with the same parameter settings as in figure 4.

References

- Alexander RD. 1987 *The biology of moral systems*. Piscataway, NJ: Aldine Transaction.
- Fehr E, Fischbacher U. 2004 Social norms and human cooperation. *Trends Cogn. Sci.* **8**, 185–190. (doi:10.1016/j.tics.2004.02.007)
- Boyd R, Richerson PJ. 2009 Culture and the evolution of human cooperation. *Phil. Trans. R. Soc. B* **364**, 3281–3288. (doi:10.1098/rstb.2009.0134)
- Henrich J *et al.* 2005 'Economic man' in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* **28**, 795–815. (doi:10.1017/S0140525X05000142)
- Ensminger J, Henrich J. 2014 *Experimenting with social norms: fairness and punishment in cross-cultural perspective*. New York, NY: Russell Sage Foundation.
- Nowak MA, Sigmund K. 1998 Evolution of indirect reciprocity by image scoring. *Nature* **393**, 573–577. (doi:10.1038/31225)
- Milinski M, Semmann D, Bakker TC, Krambeck HJ. 2001 Cooperation through indirect reciprocity: image scoring or standing strategy? *Proc. R. Soc. Lond. B* **268**, 2495–2501. (doi:10.1098/rspb.2001.1809)
- Rand DG, Nowak MA. 2013 Human cooperation. *Trends Cogn. Sci.* **17**, 413–425. (doi:10.1016/j.tics.2013.06.003)
- Wedekind C, Milinski M. 2000 Cooperation through image scoring in humans. *Science* **288**, 850–852. (doi:10.1126/science.288.5467.850)
- Bolton GE, Katok E, Ockenfels A. 2005 Cooperation among strangers with limited information about reputation. *J. Public Econ.* **89**, 1457–1468. (doi:10.1016/j.jpubeco.2004.03.008)
- Brandt H, Hauert C, Sigmund K. 2003 Punishment and reputation in spatial public goods games. *Proc. R. Soc. Lond. B* **270**, 1099–1104. (doi:10.1098/rspb.2003.2336)
- Rockenbach B, Milinski M. 2006 The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718–723. (doi:10.1038/nature05229)
- Nelissen RMA. 2008 The price you pay: cost-dependent reputation effects of altruistic punishment. *Evol. Hum. Behav.* **29**, 242–248. (doi:10.1016/j.evolhumbehav.2008.01.001)
- Chen X, Szolnoki A, Perc M. 2015 Competition and cooperation among different punishing strategies in the spatial public goods game. *Phys. Rev. E* **92**, 012819. (doi:10.1103/PhysRevE.92.012819)
- Nowak MA. 2006 Five rules for the evolution of cooperation. *Science* **314**, 1560–1563. (doi:10.1126/science.1133755)
- Capraro V, Perc M. 2021 Mathematical foundations of moral preferences. *J. R. Soc. Interface* **18**, 20200880. (doi:10.1098/rsif.2020.0880)
- Capraro V, Di Paolo R, Perc M, Pizziol V. 2024 Language-based game theory in the age of artificial intelligence. *J. R. Soc. Interface* **21**, 20230720. (doi:10.1098/rsif.2023.0720)
- Nowak MA, Sigmund K. 2005 Evolution of indirect reciprocity. *Nature* **437**, 1291–1298. (doi:10.1038/nature04131)
- Leimar O, Hammerstein P. 2001 Evolution of cooperation through indirect reciprocity. *Proc. R. Soc. Lond. B* **268**, 745–753. (doi:10.1098/rspb.2000.1573)
- Ohtsuki H, Iwasa Y. 2004 How should we define goodness?—reputation dynamics in indirect reciprocity. *J. Theor. Biol.* **231**, 107–120. (doi:10.1016/j.jtbi.2004.06.005)
- Ohtsuki H, Iwasa Y. 2006 The leading eight: social norms that can maintain cooperation by indirect reciprocity. *J. Theor. Biol.* **239**, 435–444. (doi:10.1016/j.jtbi.2005.08.008)
- Yamagishi T, Cook KS, Watabe M. 1998 Uncertainty, trust, and commitment formation in the United States and Japan. *Am. J. Sociol.* **104**, 165–194. (doi:10.1086/210005)
- Yoeli E, Hoffman M, Rand DG, Nowak MA. 2013 Powering up with indirect reciprocity in a large-scale field experiment. *Proc. Natl Acad. Sci. USA* **110**, 10424–10429. (doi:10.1073/pnas.1301210110)
- Hilbe C, Schmid L, Tkadlec J, Chatterjee K, Nowak MA. 2018 Indirect reciprocity with private, noisy, and incomplete information. *Proc. Natl Acad. Sci. USA* **115**, 12241–12246. (doi:10.1073/pnas.1810565115)
- Uchida S. 2010 Effect of private information on indirect reciprocity. *Phys. Rev. E* **82**, 036111. (doi:10.1103/PhysRevE.82.036111)
- Schmid L, Ekbatani F, Hilbe C, Chatterjee K. 2023 Quantitative assessment can stabilize indirect reciprocity under imperfect information. *Nat. Commun.* **14**, 2086. (doi:10.1038/s41467-023-37817-x)
- Kessinger TA, Tarnita CE, Plotkin JB. 2023 Evolution of norms for judging social behavior. *Proc. Natl Acad. Sci. USA* **120**, e2219480120. (doi:10.1073/pnas.2219480120)
- Kawakatsu M, Michel-Mata S, Kessinger TA, Tarnita CE, Plotkin JB. 2024 When do stereotypes undermine indirect reciprocity? *PLoS Comput. Biol.* **20**, e1011862. (doi:10.1371/journal.pcbi.1011862)
- Michel-Mata S, Kawakatsu M, Sartini J, Kessinger TA, Plotkin JB, Tarnita CE. 2024 The evolution of private reputations in information-abundant landscapes. *Nature* **634**, 883–889. (doi:10.1038/s41586-024-07977-x)
- Masuda N. 2012 Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. *J. Theor. Biol.* **311**, 8–18. (doi:10.1016/j.jtbi.2012.07.002)
- Mukherjee D, Makarius EE, Stevens CE. 2018 Business group reputation and affiliates' internationalization strategies. *J. World Bus.* **53**, 93–103. (doi:10.1016/j.jwb.2017.12.003)
- Szabó G, Hauert C. 2002 Phase transitions and volunteering in spatial public goods games. *Phys. Rev. Lett.* **89**, 118101. (doi:10.1103/PhysRevLett.89.118101)
- Szolnoki A, Perc M, Szabó G. 2009 Topology-independent impact of noise on cooperation in spatial public goods games. *Phys. Rev. E* **80**, 056109. (doi:10.1103/PhysRevE.80.056109)
- Perc M, Gómez-Gardeñes J, Szolnoki A, Floría LM, Moreno Y. 2013 Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface* **10**, 20120997. (doi:10.1098/rsif.2012.0997)
- Nax HH, Perc M, Szolnoki A, Helbing D. 2015 Stability of cooperation under image scoring in group interactions. *Sci. Rep.* **5**, 12145. (doi:10.1038/srep12145)
- Kollock P. 1998 Social dilemmas: the anatomy of cooperation. *Annu. Rev. Sociol.* **24**, 183–214. (doi:10.1146/annurev.soc.24.1.183)
- Hauert C, De Monte S, Hofbauer J, Sigmund K. 2002 Volunteering as Red Queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132. (doi:10.1126/science.1070582)
- Shao Y, Wang X, Fu F. 2019 Evolutionary dynamics of group cooperation with asymmetrical environmental feedback. *EPL* **126**, 40005. (doi:10.1209/0295-5075/126/40005)
- Wang X, Zheng Z, Fu F. 2020 Steering eco-evolutionary game dynamics with manifold control. *Proc. R. Soc. A* **476**, 20190643. (doi:10.1098/rspa.2019.0643)
- Wang X, Fu F. 2020 Eco-evolutionary dynamics with environmental feedback: cooperation in a changing world. *Europhys. Lett.* **132**, 10001. (doi:10.1209/0295-5075/132/10001)
- Jiang Y, Wang X, Liu L, Wei M, Zhao J, Zheng Z, Tang S. 2023 Nonlinear eco-evolutionary games with global environmental fluctuations and local environmental feedbacks. *PLoS Comput. Biol.* **19**, e1011269. (doi:10.1371/journal.pcbi.1011269)
- McAvoy A, Fraiman N, Hauert C, Wakeley J, Nowak MA. 2018 Public goods games in populations with fluctuating size. *Theor. Popul. Biol.* **121**, 72–84. (doi:10.1016/j.tpb.2018.01.004)
- Balliet D, Mulder LB, Van Lange PAM. 2011 Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594–615. (doi:10.1037/a0023489)
- Su Q, Wang L, Stanley HE. 2018 Understanding spatial public goods games on three-layer networks. *New J. Phys.* **20**, 103030. (doi:10.1088/1367-2630/aae6dd)
- Su Q, Li A, Wang L, Eugene Stanley H. 2019 Spatial reciprocity in the evolution of cooperation. *Proc. R. Soc. B* **286**, 20190041. (doi:10.1098/rspb.2019.0041)
- McAvoy A, Allen B, Nowak MA. 2020 Social goods dilemmas in heterogeneous societies. *Nat. Hum. Behav.* **4**, 819–831. (doi:10.1038/s41562-020-0881-2)
- Li A, Zhou L, Su Q, Cornelius SP, Liu YY, Wang L, Levin SA. 2020 Evolution of cooperation on temporal networks. *Nat. Commun.* **11**, 2259. (doi:10.1038/s41467-020-16088-w)
- Alvarez-Rodríguez U, Battiston F, de Arruda GF, Moreno Y, Perc M, Latora V. 2021 Evolutionary dynamics of higher-order interactions in social networks. *Nat. Hum. Behav.* **5**, 586–595. (doi:10.1038/s41562-020-01024-1)
- Clark D, Fudenberg D, Wolitzky A. 2020 Indirect reciprocity with simple records. *Proc. Natl Acad. Sci. USA* **117**, 11344–11349. (doi:10.1073/pnas.1921984117)

50. Suzuki S, Akiyama E. 2007 Three-person game facilitates indirect reciprocity under image scoring. *J. Theor. Biol.* **249**, 93–100. (doi:10.1016/j.jtbi.2007.07.017)
51. Panchanathan K, Boyd R. 2004 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499–502. (doi:10.1038/nature02978)
52. Perc M, Jordan JJ, Rand DG, Wang Z, Boccaletti S, Szolnoki A. 2017 Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51. (doi:10.1016/j.physrep.2017.05.004)
53. Uchida S, Sasaki T. 2013 Effect of assessment error and private information on stern-judging in indirect reciprocity. *Chaos Solitons Fractals* **56**, 175–180. (doi:10.1016/j.chaos.2013.08.006)
54. Okada I, Sasaki T, Nakai Y. 2017 Tolerant indirect reciprocity can boost social welfare through solidarity with unconditional cooperators in private monitoring. *Sci. Rep.* **7**, 9737. (doi:10.1038/s41598-017-09935-2)
55. Traulsen A, Nowak MA, Pacheco JM. 2006 Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909. (doi:10.1103/PhysRevE.74.011909)
56. Stewart AJ, Plotkin JB. 2013 From extortion to generosity, evolution in the Iterated Prisoner's Dilemma. *Proc. Natl Acad. Sci. USA* **110**, 15348–15353. (doi:10.1073/pnas.1306246110)
57. Hilbe C, Šimsa Š, Chatterjee K, Nowak MA. 2018 Evolution of cooperation in stochastic games. *Nature* **559**, 246–249. (doi:10.1038/s41586-018-0277-x)
58. Szabó G, Töke C. 1998 Evolutionary prisoner's dilemma game on a square lattice. *Phys. Rev. E* **58**, 69–73. (doi:10.1103/PhysRevE.58.69)
59. Fudenberg D, Imhof LA. 2006 Imitation processes with small mutations. *J. Econ. Theory* **131**, 251–262. (doi:10.1016/j.jet.2005.04.006)
60. Wu B, Gokhale CS, Wang L, Traulsen A. 2012 How small are small mutation rates? *J. Math. Biol.* **64**, 803–827. (doi:10.1007/s00285-011-0430-8)
61. Wuchty S, Jones BF, Uzzi B. 2007 The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039. (doi:10.1126/science.1136099)
62. Wu L, Wang D, Evans JA. 2019 Large teams develop and small teams disrupt science and technology. *Nature* **566**, 378–382. (doi:10.1038/s41586-019-0941-9)
63. Gelfand MJ, Nishii LH, Raver JL. 2006 On the nature and importance of cultural tightness-looseness. *J. Appl. Psychol.* **91**, 1225–1244. (doi:10.1037/0021-9010.91.6.1225)
64. Gelfand MJ *et al.* 2011 Differences between tight and loose cultures: a 33-nation study. *Science* **332**, 1100–1104. (doi:10.1126/science.1197754)
65. Harrington JR, Gelfand MJ. 2014 Tightness-looseness across the 50 United States. *Proc. Natl Acad. Sci. USA* **111**, 7990–7995. (doi:10.1073/pnas.1317937111)
66. Stewart AJ, Raihani N. 2023 Group reciprocity and the evolution of stereotyping. *Proc. R. Soc. B* **290**, 20221834. (doi:10.1098/rspb.2022.1834)
67. Dovidio JF, Gaertner SL. 1993 Stereotypes and evaluative intergroup bias. In *Affect, cognition and stereotyping*, pp. 167–193. San Diego, CA: Academic Press. (doi:10.1016/B978-0-08-088579-7.50012-0)
68. Schuster P, Sigmund K. 1983 Replicator dynamics. *J. Theor. Biol.* **100**, 533–538. (doi:10.1016/0022-5193(83)90445-9)
69. Roca CP, Cuesta JA, Sánchez A. 2009 Evolutionary game theory: temporal and spatial effects beyond replicator dynamics. *Phys. Life Rev.* **6**, 208–249. (doi:10.1016/j.plev.2009.08.001)
70. RoyWey. 1998 Indirect-reciprocity-with-collective-reputation. GitHub. See <https://github.com/RoyWey1998/Indirect-reciprocity-with-collective-reputation>.
71. Wei M, Wang X, Liu L, Zheng H, Jiang Y, Hao Y *et al.* 2025 Supplementary material from: Indirect reciprocity in the public goods game with collective reputations. Figshare. (doi:10.6084/m9.figshare.c.7669668)