Evolution of cooperation under indirect reciprocity and arbitrary exploration rates

Fernando P. Santos\(^1,2\), Jorge M. Pacheco\(^2,3,4\) & Francisco C. Santos\(^1,2\)

Cooperation has been recognized as an evolutionary puzzle since Darwin, and remains identified as one of the biggest challenges of the XXIst century. Indirect Reciprocity (IR), a key mechanism that humans employ to cooperate with each other, establishes that individual behaviour depends on reputations, which in turn evolve depending on social norms that classify behaviours as good or bad. While it is well known that different social norms give rise to distinct cooperation levels, it remains unclear how the performance of each norm is influenced by the random exploration of new behaviours, often a key component of social dynamics where a plethora of stimuli may compel individuals to deviate from pre-defined behaviours. Here we study, for the first time, the impact of varying degrees of exploration rates – the likelihood of spontaneously adopting another strategy, akin to a mutation probability in evolutionary dynamics – in the emergence of cooperation under IR. We show that high exploration rates may either improve or harm cooperation, depending on the underlying social norm at work. Regarding some of the most popular social norms studied to date, we find that cooperation under Simple-standing and Image-score is enhanced by high exploration rates, whereas the opposite occurs for Stern-judging and Shunning.

The act of cooperation is generally framed quantitatively as an interaction in which an individual provides a benefit $b$ to another at a cost $c$ to himself$^1-3$. Traditionally, one assumes that the benefit exceeds the cost ($b > c$). This means that whenever two individuals are given the option to cooperate or not (that is, to defect) with each other, the joint social optimum is achieved when both cooperate; yet, for each one individually, the fact that cooperation is costly configures defection as the preferred option$^4$. In this sense, cooperation embodies a fascinating social dilemma within societies. When played bilaterally and simultaneously, this donation game turns into a Prisoner’s Dilemma (PD), which provides a convenient abstraction of a wide range of human interactions where the pursuit of self-interest leads to poor collective outcomes. In this context, two fundamental questions have puzzled scientists since Darwin: How is cooperation so widespread in human societies? How can cooperation emerge where it is absent?

The human capacity to establish and use reputation systems suggests some answers. Indeed, humans developed a huge machinery used at profit to share information about others$^5,6$; they evolved to shape decision-making based on the reputation of those they interact with; and they act influenced by what they want others to know about them. All together, reputations work as a mechanism of social control and a lever for cooperation, altruism and collective action$^7-10$.

Cooperation and reputation have been mathematically linked in models of Indirect Reciprocity (IR)$^{11-42}$. IR models comprise individuals who adopt heuristics for decision-making based on the reputation of those they interact with; and they act influenced by what they want others to know about them. All together, reputations work as a mechanism of social control and a lever for cooperation, altruism and collective action$^7-10$.

Cooperation and reputation have been mathematically linked in models of Indirect Reciprocity (IR)$^{11-42}$. IR models comprise individuals who adopt heuristics for decision-making based on reputations. In general, the complexity of IR models is limitless. Let us consider the simplest scenario where reputations can either be good (G) or bad (B), and individuals can adopt one of two possible actions: to cooperate (C) or to defect (D). As before, if they cooperate they loose $c$ and the opponent earns $b$. Otherwise, they pay no cost and confer no benefit. The decision to opt between C and D is not arbitrary. It is encoded in an action rule that prescribes what to do against a G or a B opponent. Naturally, four strategies emerge: always cooperate (AllC); always defect (AllD); discriminate between G and B, only cooperating with G (Disc); or cooperate uniquely with a B (pDisc). Here we assume that

\(^1\)INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, IST-Taguspark, 2744-016 Porto Salvo, Portugal. \(^2\)ATP-group, P-2744-016 Porto Salvo, Portugal. \(^3\)Centro de Biologia Molecular e Ambiental, Universidade do Minho, 4710 - 057 Braga, Portugal. \(^4\)Departamento de Matemática e Aplicações, Universidade do Minho, 4710 - 057 Braga, Portugal. Correspondence and requests for materials should be addressed to F.C.S. (email: franciscocsantos@tecnico.ulisboa.pt)
the adoption of each strategy follows a process of social learning\textsuperscript{43} i.e., at each time step, one individual is picked and is given the opportunity to imitate the strategy of a model agent depending on its (better) performance.

Within this environment, reputations are dynamic. After each interaction, the reputation of the individual who decided between C or D (called the donor) when facing another individual (called the recipient) will eventually change. The rule that settles the new reputation to the donor (innermost ring with G = Blue and B = Grey) and on the action of the donor (middle ring, where C = Orange and D = Black). There are 16 possible social norms combining all these bits of information. 4 of the most popular are (A) Stern-judging: [G, B, B, G]; (B) Shunning: [G, B, B, B]; (C) Image-score: [G, B, G, B] and (D) Simple-standing: [G, B, G, G].

Figure 1. Ring representation of 4 of the most popular social norms. A social norm can be thought of as a vector of the type \([w, x, y, z]\). Each position provides the new reputation of the donor and, in a ring illustration, this information is placed in the outer ring with G = Blue and B = Grey (panel A). This new reputation depends on the reputation of the recipient (innermost ring with G = Blue and B = Grey) and on the action of the donor (middle ring, where C = Orange and D = Black). There are 16 possible social norms combining all these bits of information. 4 of the most popular are (A) Stern-judging: [G, B, B, G]; (B) Shunning: [G, B, B, B]; (C) Image-score: [G, B, G, B] and (D) Simple-standing: [G, B, G, G].
We resort to the toolkit of evolutionary game theory (EGT) and numerically explore the dynamics of strategy per state. Social learning — and thereby reputations and social norms — play the steering role in strategy adoption. Those norms that already amplify the levels of cooperation under a regime of social learning, as is the case of cooperation under social learning benefit from the co-existence of all strategies. Naturally, the converse happens for those social norms that are unable to promote cooperation. In this work, we address this issue in the context of IR. We resort to the toolkit of evolutionary game theory (EGT) and numerically explore the dynamics of strategy adoption, when social norms govern the co-evolving dynamics of reputation assignment and when individuals may spontaneously adopt (with arbitrary exploration rates) any strategy. We compute the so-called stationary distribution of strategies in finite populations of size Z, and the population-wide gradients of selection, which allow us to characterize, in detail, the dynamics of strategy adoption. We find that the strategy ambiguity stemming from high exploration rates favours cooperation under the norms Image-Score (IS) and Simple-Standing (SS), whereas it inhibits cooperation under Shunning (SH) and Stern-Judging (SJ).

**Results**

In Fig. 2 we depict, for a wide interval of exploration rates (from $10^{-3}/Z$ to 1) and $Z = 50$, the cooperation levels associated with each social norm. We use specific colours to represent the behaviour associated with the 4 most popular order social norms, defined in Fig. 1. As shown, whenever $\mu > 10^{-3}/Z$ (indicated by the leftmost vertical dashed line) the cooperation level associated with each social norm is clearly affected by high exploration rates, with the ranking of each norm even changing. While SJ and SS preserve the status of norms that noticeably sustain more cooperation, there are important effects worth pointing for large values of $\mu$: (1) IS and SS benefit, in most cases, from higher exploration rates; (2) the cooperation rate under SH and SJ slightly decreases for high exploration rates and (3) as $\mu$ approaches 1, cooperation in all the remaining social norms (drawn with a black colour) generally increases with $\mu$, while cooperation rates under all norms approach 0.5.

Let us start by clarifying point 3): Increasing the exploration rate implies that social learning plays a decreasing role in the overall evolution of decision making. In particular, for very large mutation rates, since we have a significant fraction of the population ($\mu$) just exploring the strategy space, the system typically does not access the grey shaded areas pictured in Fig. 2 (see Methods for details). This way, those social norms that are unable to promote cooperation under social learning benefit from the co-existence of all strategies. Naturally, the converse happens for those norms that already amplify the levels of cooperation under a regime of social learning, as is the case of SJ and SS. In the following discussion, we focus on points (1) and (2), i.e., on the interval of $\mu$ in which fitness dependent social learning — and thereby reputations and social norms — play the steering role in strategy adoption.

To clarify the points 1) and 2) and to further understand the effects of high exploration rates on norms SH, SJ, IS, and SS, we show in Figs 3 and 4 i) the gradients of selection ($\Gamma_{\mu}$), together with ii) the reputation distribution per state ($\gamma_{\mu}$) and iii) the stationary distribution ($\lambda_{\mu}$) of strategies, for the 4 most popular social norms (see section Methods for a detailed account of these metrics). The tetrahedrons represent the entire state space (simplex) defined by this 4-strategy dynamics, where each corner defines a monomorphic state in which the entire population adopts the same strategy. For simplicity and visualization purposes we provide details of the evolutionary dynamics inside the triangular slices of the complete 3D dynamics. Inside these triangular slices, each arrow corresponds to $\Gamma_{\mu}$ and represents the most probable direction of evolution. The prevalence in each state ($\lambda_{\mu}$) is associated with the background colour intensity whereas blue/red tones translate into an increased number of G/B individuals ($\gamma_{\mu}$).

**Figure 2.** Effect of exploration rate in the cooperation sustained by each social norm. Cooperation under Simple-Standing (SS) and Image-score (IS) profits from high exploration rates ($\mu$); conversely, cooperation rates under shunning (SH) and Stern-judging (SJ) decrease with high $\mu$ (see Fig. 1 for definitions of these social norms). The other social norms considered are unable to promote cooperation under a wide range of $\mu$ (black lines). When $\mu$ approaches 1, individuals pick strategies randomly and social learning has no effect in the overall evolution of decision making; hence cooperation increases compared with the reference scenario. Furthermore, for high $\mu$ there is a configuration subspace that becomes inaccessible (i.e., $\mu$ implies a minimum prevalence of each strategy), which imposes limits to cooperation. These limits are represented with a grey background. Dashed lines represent the exploration rates studied in detail in Figs 3 and 4. Other parameters (see section Methods): $Z = 50$, $b = 5$, $c = 1$, $\chi = \epsilon = \alpha = 0.01$, $\tau = 1$. 

sort of individual or social influence. This process interestingly resembles (and may be formulated in a mathematically identical way) biological mutations in genetic settings. However, whereas in genetics random mutations are often rare, in social evolution this is not necessarily the case. On the contrary, high exploration rates may turn out to be the norm rather than the exception, which may strongly affect the evolutionary dynamics of populations facing cooperation and fairness dilemmas. In this work we address this issue in the context of IR.
Figure 3 shows that the harshness of SH becomes more evident for high $\mu$ (Fig. 3-B), and cooperation is thereby precluded. High $\mu$ means that B labels are even easier to be attributed due to the presence of individuals that spontaneously adopt AllD and pDisc, strategies that in general are labelled B under SH. The increase in B individuals makes Disc and AllD almost indistinguishable and turns the coordination between these strategies (Fig. 3-A) into a co-existence dynamics (Fig. 3-B). In the case of SJ we observe that, for low $\mu$, most of the time is spent in the highly cooperative monomorphic states Disc and pDisc (Fig. 3-C,D). When $\mu$ increases (Fig. 3-E) cooperation declines as the prevalence under monomorphic states Disc/pDisc decreases, with individuals exploring the other strategies (AllD and AllC).

Figure 4 reveals why cooperation increases at high $\mu$, when the prevailing norm is either IS or SS. For IS, high exploration rates have the ability to move the system away from states where AllD is the prevalent strategy (Fig. 4-B). Indeed, high values of $\mu$ increase the number of individuals that spontaneously adopt Disc or AllC, placing and stabilizing the dynamics in the interior of the simplex. In the case of SS, high values of $\mu$ allow the population to overcome the “coordination barrier” between AllD and Disc (Fig. 4-E), thus making it easier to achieve the minimum number of Discs that renders advantageous to have a G reputation; this way, the population spends less time in AllD and more time near the edge where AllC and Disc co-exist, which naturally benefits cooperation.

Besides the particular effect of $\mu$ in each social norm, it is noteworthy that cooperation levels remain qualitatively unchanged (apart from the slow monotonic increase of cooperation under SS) for a wide range of values of the exploration rate $\mu$ (Fig. 2). These results confirm, for the first time, that for a wide interval of values of $\mu$ the Small-Mutation Approximation (SMA) proves accurate in the context of IR. With SMA, one assumes that $\mu$ is small enough so that the system will spend a negligible fraction of time in polymorphic states. This way, SMA allows a convenient characterization of evolutionary processes, formally in the limit when $\mu \ll 1/Z$, through a
reduced (embedded) Markov chain involving monomorphic states only. Yet, its range of applicability is often unclear\(^4^8\), requiring the use of large-scale computer simulations as the ones performed here.

**Discussion**

In this work we investigate the effect of an arbitrary exploration rate on the evolutionary dynamics of cooperation under indirect reciprocity. This is an analysis of general interest given the ubiquity of creative and emotional traits that characterize human behaviour, urging them to explore new strategies\(^4^7\). The reputation systems studied here were based on the action of the donor and the reputation of the recipient, a feature associated with so-called 2\(^{nd}\)-order norms\(^2^1,^3^1\). We show that random exploration of the strategy space has a non-trivial effect in the behaviour dynamics of populations under IR: when high, it increases cooperation under IS (interestingly, a norm just demanding 1\(^{st}\) order information) and SS, while it decreases cooperation under SH and SJ. Despite these key observations, our results indicate that the so-called “leading two” norms\(^2^1\) — SJ and SS — remain the most effective in promoting cooperation, a result that may not hold in the space of high-order norms. Overall, our results suggest that a general heuristics can, in fact, be intuitively derived: when the social norm is able to provide high cooperation rates by relying on the stability of some cooperative monomorphic states (e.g., Disc in SJ and SH), exploration is pernicious; when the social norm is unable to provide that stability, allowing instead for the occurrence of some uncooperative monomorphic state (e.g., AllD in IS and in SS) high exploration rates favour overall cooperation by allowing the population to move away from this undesirable state.

Despite these results, we confirm that SJ remains the social norm that, overall, promotes the highest levels of cooperation. The strength of this norm relies on the efficiency to promote a strong stability of two highly cooperative monomorphic states. No other polymorphic state (i.e., in which more than one strategy co-exists) is able to reach those levels of cooperation; hence, any ambiguity source that moves the population to the interior of the simplex will be detrimental to cooperation. This, however, is not the whole story, since the reference levels of cooperation attained under SJ are so high that the disadvantageous effect induced by large exploration rates does not prevent SJ to remain the leading norm in what concerns the promotion of cooperation.

The dynamics under SJ also puts in evidence a remarkable fact. Because of the inherent symmetry of SJ with respect to the G and B labels, the net effect of SJ resembles some sort of a divide-and-conquer procedure. Indeed,
under \(SJ\) the whole state space is divided into two smaller basins of attraction that push the population to full cooperation in both cases: if there is a majority of individuals that cooperate with B and defect with G (pDisc), the population will move into a state where everyone is B; if there is a majority of Disc, the population will move to a state where everyone is G; in both limits, cooperation prevails, and that ensures the collective optimum (Fig. 3-D).

Ultimately, the notions of Good and Bad, as we normally use them, are defined by the actions and not by the labels attributed to specific reputations. Clearly, in the case of \(SJ\), the meaning of the signals G(ood) or B(ad) may emerge from a simple convention\(^{32,53}\). This means, in turn, that if we would consider different populations evolving independently and under the assessment of this norm, full cooperation might still be achieved under different emergent conventions for what is Good or Bad. Given the conceptual simplicity of the underlying IR model employed, this constitutes a remarkable feature of \(SJ\), related also to complex topics such as conflicting moral systems\(^{30,34}\) or the appearance of in/out groups together with their inherent normative values. These observations can only be made, however, to the extent that (like we have done in this work), one provides the global dynamics considering the 4 possible action rules (AllD, Disc, pDisc, AllC), instead of carrying out the analysis including only 3, as often happens. This feature, together with the systematic investigation of high exploration rates in IR, were carried out here for the first time.

It is also noteworthy the way \(SJ\) implies a moral judgement that, besides justifying the defection against undesirable opponents, also condemns cooperating with those. Indeed, under \(SJ\), whoever cooperates with an opponent carrying a Bad reputation gets himself a Bad reputation. Interestingly, this judgement is somehow verified in the behaviour of toddlers, who prefer those that mis treat (rather than help) opponents that misbehaved in the past\(^{55–58}\).

Finally, given the current importance of reputation-based systems, indirect reciprocity models emerge nowadays as relevant toolkits for artificial intelligence applications, web platforms and systems supporting sharing economies. Online communities are nowadays pervasive and most of them profit from the readiness of its users to cooperate, which is often supported by reputation\(^{49}\). Overall, reputation mechanisms are considered a key element in the design of multiagent systems\(^{60–62}\). The simulation technique that we present here, together with the proposed metrics to visualize the emergent dynamics, can provide the appropriate basis in which to study (and evaluate ref. 63) other challenges related with strategic dynamics and reputation systems, such as the effect of considering different reputation management schemes\(^{64–67}\), the design of new underlying structures of interaction\(^{36,68–74}\) or the formalization of bottom-up artificial morality and machine ethics\(^{75}\).

**Methods**

**Strategies and reputations.** We model a population of \(Z\) individuals that interact with each other and may change their strategy over time (in the way described below). Individuals play with each other a donation game that reproduces, in a simple way, the mathematics of cooperation\(^{45}\). Each individual may choose one of two actions: to cooperate (C), meaning that the donor incurs a cost \(c\) to confer a benefit \(b\) to the opponent; to defect (D), whereby neither player earns or looses anything. Each individual follows a strategy (action rule) that defines how to act against a Good (G) or Bad (B) opponent. This rule is encoded in a vector \([x, y]\) where the first position dictates what to do (C/D) against a G and the second position what to do against a B. There are four possible strategies: AllC \([C, C]\), Disc \([C, D]\), pDisc \([D, C]\) and AllD \([D, D]\). Initially, each individual has one reputation (G or B) and strategy; these are attributed randomly with uniform probability.

**Update of strategies and reputations.** At each (discrete) time step \((t)\) one individual \(x\) is randomly selected to update its strategy. With probability \(\mu\) (so called exploration rate\(^{61,66,48}\)) \(x\) adopts a random strategy within the full space of possible strategies. With probability \(1-\mu\) strategy change may take place through social learning; in this case, \(x\) compares its fitness \((f_x)\) with that \((f_y)\) of another individual \(y\) randomly selected, changing its strategy to that of \(y\) with a probability that increases with the fitness difference, given by \(p_{xy} = (1 + e^{-\beta(f_y-f_x)})^{-1}\), with \(\beta=1\), ensuring a significant selection strength. This imitation process and the associated probability function are well documented\(^{72,73}\) and known as pairwise comparison rule. Both fitnesses \(f_x\) and \(f_y\) are associated with the average payoff obtained in 2Z donation games, always played against random opponents. In each of these games, both individuals play once as donor and once as recipient. After each donation game, with a probability \(\tau\), a new reputation is attributed to the individual acting as donor, in accordance with the social norm fixed in the population. With probability \(1-\tau\), the donor keeps the same reputation. For simplicity, we assume a public reputations scheme\(^{6,19,20,31}\) such that, through gossip or rumours, new reputations spread to everyone, a simplification that can naturally be relaxed in future works based on different reputation database schemes\(^{64}\). As already described in the Introduction and depicted in Fig. 1, here we consider that a social norm is a vector of the type \([w, x, y, z]\), where each position provides the information regarding the new reputation of the donor (G/B), given the reputation of the recipient (G/B) and the action of the donor (C/D).

**Errors.** We allow for the inclusion of errors of three different types: execution errors – with probability \(\varepsilon\), there is a failure to cooperate when the action rule dictates so\(^{34}\); assignment errors – with probability \(\alpha\), the assigned reputation is the opposite of the one prescribed by the social norm; and private assessment errors – with probability \(\chi\), when deciding about what action to employ or when deciding the next reputation of an opponent, the retrieved reputation is the opposite of the one actually owned\(^{31}\). The effect of each kind of error in the cooperation levels supported by each specific social norm was already discussed\(^{31}\). While Figs 1, 2 and 3 report the results for \(\varepsilon = \alpha = \chi = 0.01\), we confirm that the effect of exploration rate discussed in the main text remains unchanged if (all else constant) \(\varepsilon, \alpha\) or \(\chi\) go up to 0.1 and \(\tau\) down to 0.01.
Tracing the dynamics. Each possible state of this population can be enumerated and arranged spatially, so that its dynamics becomes straightforward to visualize. As there are Z individuals and 4 different strategies, each state of the population is identified by a tuple \( x = (k_0, k_1, k_2) \) meaning that, in state \( x \), there are \( k_0 \) individuals adopting strategy AllD, \( k_1 \) adopting strategy AllC, \( k_2 \) adopting strategy Disc, and \( k_3 = Z - k_0 - k_1 - k_2 \) adopting strategy AllC. In total, there are \((Z + 1)(Z + 2)(Z + 3)/6\) states. These states can intuitively be assembled in a 3D simplex (a tetrahedron as in Figs 3 and 4), where the vertices represent states in which all the agents are adopting the same one unique strategy (monomorphic or pure states). In Figs 3 and 4, to allow for an easier interpretation of the results, we present the dynamics along cross sections of this space (2D simplexes, triangles of variable size depending on the specific cross section position). Provided this description, we retrieve information in four different forms discussed below: average cooperation rates, average time in each state, average fraction of good and bad reputations, and gradient of selection.

Average cooperation rate. We use the general result that one generation corresponds to \( Z \) discrete time steps where a strategy update may occur, and we simulate the system for \( G = 10^7 \) generations. We repeat each simulation \( R = 100 \) times (runs), such that in each run the pseudo-random number sequence will be different. The average cooperation rate \( \langle \eta \rangle \) in run \( i \) is computed by dividing the total number of cooperative acts \( (C_i) \) by the total number of donation games \( (K_i) \):

\[
\eta = R^{-1} \sum_{i=1}^{R} C_i/K_i = R^{-1} \sum_{i=1}^{R} \eta_i
\]

As \( \mu \to 1 \), exploration increasingly dominates selection (in the form of social learning), implying that reputations and social norms are subjected to minor perturbations with respect to a purely random choice process. Particularly, since there is always the probability \( \mu \) that an individual explores any other strategy, the space of accessible configurations is effectively reduced with the increase of \( \mu \), such that \( \eta \in [\eta_{\text{min}}, \eta_{\text{max}}] \). The minimum fraction \( \eta_{\text{min}} \equiv \eta_{\text{min}}(x)/Z \) is the fraction of cooperation \( C_i \) by strategy \( i \) is given by one of the roots of \( \gamma_i(x) = T_i(x) - T_i'(x) \), where \( T_i'(x) = (1 - \mu)x \sum_{j \neq i} x_j + \mu \sum_{j \neq i} x_j \) and \( T_i(x) = (1 - \mu)x \sum_{j \neq i} x_j + \mu \sum_{j \neq i} x_j \) are the probabilities of increasing and decreasing (respectively) the number of individuals of strategy \( i \), for a general state \( x \).

In the most unfavourable scenario for strategy \( i \) (i.e., strong selection \( \beta \gg 1 \), \( p_{i \to i} = 0 \) and \( p_{i \to j} = 1 \)), \( \eta_{\text{min}} = x \) can be calculated by solving \( \mu(1 - xd)/d - (1 - \mu)x(1 - x) = 0 \), which results in \( x = [2(1 - \mu)]^{-1}(1 - \mu + \mu - 1/d) \). This quantity is close to \( \mu/d \) as employed in ref. 47. In general \( \eta = k_i/Z + \gamma k_i/Z + (1 - \gamma)k_i/Z \) with \( \gamma \) standing for the fraction of G individuals (see below) — which is minimized for \( k_i/Z + k_i/Z = k_i/Z + x_i \) such that, for \( d = 4 \), \( \eta_{\text{min}} = x + \gamma x + (1 - \gamma) x = 2x \). A similar argument can be used to compute the minimum defection rate, leading to a \( \eta_{\text{max}} = 1 - \eta_{\text{min}} \). The inaccessible cooperation rates \( \eta \notin [\eta_{\text{min}}, \eta_{\text{max}}] \) are represented in Fig. 1 by a grey background.

Average time in each state. We keep a counter that totalizes the number of time steps a given state of the population is reached. After the proper normalization (the total number of steps is GRZ), we collect information on the average fraction of time \( \lambda_i \) that the population spends in each state \( \lambda_i \). This information is conveyed in the triangles of Figs 3 and 4 by means of colour intensity. In the tetrahedrons, we place a sphere in the simplex (a tetrahedron as in Figs 3 and 4), where the vertices represent states in which all the agents are adopting a unique strategy. In Figs 3 and 4, to allow for an easier interpretation of the results, we present the dynamics along cross sections of this space (2D simplexes, triangles of variable size depending on the specific cross section position). Provided this description, we retrieve information in four different forms discussed below: average cooperation rates, average time in each state, average fraction of good and bad reputations, and gradient of selection.

Average fraction of Good and Bad reputations. By directly placing the system in each possible state \( x = (k_0, k_1, k_2) \) and simulating one time step per run (total number of runs \( R \), each starting in the same state \( x \)), we can draw a picture of the average number of G/B. After each time step we save the resulting fraction of G individuals, \( G_{b,x}/(G_{b,x} + B_{b,x}) \). After all runs, we have the average fraction of G individuals accruing from a specific state \( x \):

\[
\gamma_x = R^{-1} \sum_{i=1}^{R} G_{b,x}/(G_{b,x} + B_{b,x})
\]

Gradient of selection. Following a similar procedure as the one described in the previous section, we can compute the transition probability between each pair of neighbour states \( x \) and \( y \). For each state \( x = (k_0, k_1, k_2) \), we count the number of times that, after each time step, the system moves into each of the (maximum) 12 adjacent states \( y = (l_0, l_1, l_2) \). For that, we keep the quantities \( \beta_{x,y} \) and \( \delta_{x,y} \), i.e., the number of times that a strategy \( i \) is born \( (l_i = k_i + 1) \) or dies \( (l_i = k_i - 1) \) in state \( x \). We keep the number of runs that we start a time step in each state \( x \), \( R_x \). This provides the required information to represent the so-called gradient of selection \( \Gamma \), i.e., a vector field that provides an approximation for the most probable direction that the system will follow once located in state \( x \), given the update of strategies (imitation and exploration) by each agent (described above),

\[
\Gamma_x = (\beta_{x,y} - \delta_{x,y})/R_x
\]

References
1. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. Science 211, 1390–1396 (1981).
2. Rand, D. G. & Nowak, M. A. Human cooperation. Trends. Cogn. Sci. 17, 413–425 (2013).
3. Nowak, M. A. Five rules for the evolution of cooperation. Science 314, 1560–1563 (2006).
4. Dawes, R. M. Social dilemmas. Annu. Rev. Psychol. 31, 169–193 (1980).
41. Nax, H. H., Perc, M., Szolnoki, A. & Helbing, D. Stability of cooperation under image scoring in group interactions. *J. Theor. Biol.* 239, 183–194 (2006).

39. Fu, F., Hauert, C., Nowak, M. A. & Wang, L. Reputation-based partner choice promotes cooperation in social networks. *Biol. Lett.* 2, 117–120 (2006).

37. Uchida, S. & Sigmund, K. The competition of assessment rules for indirect reciprocity. *J. Theor. Biol.* 415, 359–363 (2002).

35. Leimar, O. & Hammerstein, P. Evolution of cooperation through indirect reciprocity. *Nature* 393, 357–358 (1998).

34. Fishman, M. A. Indirect reciprocity among imperfect individuals. *Proc. R. Soc. Lond. B* 271, 221–228 (2004).

32. Ohtsuki, H., Iwasa, Y. & Nowak, M. A. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *J. Theor. Biol.* 245, 539–552 (2007).

20. Ohtsuki, H. & Iwasa, Y. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *J. Theor. Biol.* 244, 518–531 (2007).

17. Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *Nature* 415, 424–426 (2002).

15. Ghang, W. & Nowak, M. A. Indirect reciprocity with optional interactions. *Proc. Natl. Acad. Sci. USA* 107, 2693–2696 (2010).

13. Brandt, H. & Sigmund, K. The good, the bad and the discriminator—errors in direct and indirect reciprocity. *Nature* 415, 421–423 (2002).

56. Hamlin, J. K. & Wynn, K. Youn infants prefer prosocial to antisocial others. *Proc. Natl. Acad. Sci. USA* 107, 8971–8975 (2010).

55. Tasimi, A. & Wynn, K. Costly rejection of wrongdoers by infants and children. *Cognition* 151, 76–79 (2016).

54. Greene, J. *Moral Tribes: Emotion, Reason and the Gap Between Us and Them.* (Atlantic Books, 2013).
59. Jøsang, A., Ismail, R. & Boyd, C. A survey of trust and reputation systems for online service provision. *Decis. Support. Syst.* **43**, 618–644 (2007).
60. Pinyol, J. & Sabater-Mir, J. Computational trust and reputation models for open multi-agent systems: a review. *Artif. Intell. Rev.* **40**, 1–25 (2013).
61. Conte, R. & Paolucci, M. *Reputation in Artificial Societies: Social Beliefs for Social Order* Vol. 6 (Springer Science & Business Media, 2002).
62. Castelfranchi, C. & Falcone, R. In *Proceedings of the 1998 International Conference on Multi Agent Systems* 72–79 (IEEE, 1998).
63. Hazard, C. I. & Singh, M. P. In *Proceedings of the International Joint Conference Artificial Intelligence, IJCAI 2013* 191–197 (AAAI Press, 2013).
64. Bachrach, Y., Parunak, H. V., Proctor, M. & Toueg, S. Gossip-based aggregation of trust in decentralized reputation systems. *Auton. Agent Multi Agent Syst.** **19**, 153–172 (2009).
65. Chau, W. Y. C., Weng, J. & Shi, S. K. Direct vs. indirect reciprocity: an experimental study. *J. Econ. Behav. Organ.* **75**, 129–142 (2010).
66. Takeda, S., Suzuki, H. & Masuda, N. Indirect reciprocity with trinary reputations. *J Theor Biol.* **317**, 338–347 (2013).
67. Sasaki, T., Okada, I. & Nakai, Y. The evolution of conditional moral assessment in indirect reciprocity. *arXiv preprint arXiv:1605.02166* (2016).
68. Sabater, J. & Sierra, C. Review on computational trust and reputation models. *Artif. Intell. Rev.* **24**, 33–60 (2005).
69. Pinheiro, F. L., Santos, M. D., Santos, F. C. & Pacheco, J. M. Origin of peer influence in social networks. *Phys. Rev. Lett.* **112**, 098702 (2014).
70. Santos, F. C. & Pacheco, J. M. Scale-free networks provide a unifying framework for the emergence of cooperation. *Phys. Rev. Lett.* **95**, 098104 (2005).
71. Santos, F. C., Santos, M. D. & Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature* **454**, 213–216 (2008).
72. Szabó, G. & Fath, G. Evolutionary games on graphs. *Phys Rep.* **446**, 97–216 (2007).
73. Perc, M. & Szolnoki, A. Coevolutionary games—a mini review. *BioSystems* **99**, 109–125 (2010).
74. Wang, Z., Wang, L., Szolnoki, A. & Perc, M. Evolutionary games on multilayer networks: a colloquium. *Eur. Phys. J. B* **88**, 1–15 (2015).
75. Pereira, L. M. & Saptawijaya, A. Bridging two realms of machine ethics. *Rethinking Machine Ethics in the Age of Ubiquitous Technology* 197–224 (2015).
76. Szabó, G. & Tóke, C. Evolutionary prisoner’s dilemma game on a square lattice. *Phys. Rev. E* **58**, 69 (1998).

**Acknowledgements**
The authors thank Vítor V. Vasconcelos for fruitful discussions. This research was supported by Fundação para a Ciência e Tecnologia (FCT) through grants SFRH/BD/94736/2013, PTDC/EEI-SII/5081/2014, PTDC/MAT/STA/3358/2014 and by multiannual funding of CBMA and INESC-ID (under the projects UID/BIA/04050/2013 and UID/CEC/59021/2013) provided by FCT.

**Author Contributions**
F.P.S., J.M.P. and F.C.S. designed and implemented the research; F.P.S., J.M.P. and F.C.S. prepared all the Figures; F.P.S., J.M.P. and F.C.S. wrote the manuscript; F.P.S., J.M.P. and F.C.S. reviewed the manuscript.

**Additional Information**
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Santos, F. P. et al. Evolution of cooperation under indirect reciprocity and arbitrary exploration rates. *Sci. Rep.* **6**, 37517; doi: 10.1038/srep37517 (2016).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016