# Modelling stereotyping in cooperation systems

Wafi Bedewi[1], Roger M. Whitaker[1], Gualtiero B. Colombo[1], Stuart M Allen[1], and Yarrow Dunham[2]

[1] School of Computer Science & Informatics, Cardiff University, Cardiff, UK
[2] Department of Psychology, Yale University, New Haven, CT 06511, USA
{BedewiWA,WhitakerRM,ColomboG,AllenSM}@cardiff.ac.uk
yarrow.dunham@yale.edu

**Abstract.** Cooperation is a sophisticated example of collective intelligence. This is particularly the case for indirect reciprocity, where benefit is provided to others without a guarantee of a future return. This is becoming increasingly relevant to future technology, where autonomous machines face cooperative dilemmas. This paper addresses the problem of stereotyping, where traits belonging to an individual are used as proxy when assessing their reputation. This is a cognitive heuristic that humans frequently use to avoid deliberation, but can lead to negative societal implications such as discrimination. It is feasible that machines could be equally susceptible. Our contribution concerns a new and general framework to examine how stereotyping affects the reputation of agents engaging in indirect reciprocity. The framework is flexible and focuses on how reputations are shared. This offers the opportunity to assess the interplay between the sharing of traits and the cost, in terms of reduced cooperation, through opportunities for shirkers to benefit. This is demonstrated using a number of key scenarios. In particular, the results show that cooperation is sensitive to the structure of reputation sharing between individuals.

**Keywords:** stereotyping, cooperation, indirect reciprocity, reputation

## 1 Introduction

Cooperation is a sophisticated form of collective intelligence where individuals become incentivised to help one another and benefit from a coalition. One particularly interesting but challenging form of cooperation is *indirect reciprocity*, which is complex because it involves donating to a third party without any guarantee of future reciprocation. Cooperation in this form involves a small cost to the donor, and a much larger benefit to the recipient. This is a hallmark of human behaviour that leads to a societal benefit, by providing a resource through which unrelated individuals support each other [1, 3].

Extensive research has been successful in establishing conditions and mechanisms that promote indirect reciprocity. However, as machines are developed that feature cognition and autonomy, interest in cooperation is reaching beyond humans [20]. Transportation is just one emerging example where technology,

through autonomous vehicles, will encounter cooperative decision making [14]. This scenario features latent indirect reciprocity, such as when one driver allows another to manoeuvre in traffic. Journeys in congestion often depend on this, such as when exiting a T-junction, without which safe progress would be impossible in many cities.

Beyond technological scenarios, persistent human scenarios such as intergroup conflict [37] continue to motivate the exploration of cooperation, and the basis for it being sustained. The decision on whether or not to cooperate, when called upon, is the fundamental issue. *Reputation* is an important component that provides a currency through which cooperation can be recognised and signalled [24], allowing individuals to leverage future help when needed [22]. In recent times reputation systems have also emerged to support decision making in diverse areas of e-commerce [19, 31, 40] for example. There are also many areas of work in multi-agent systems where the focus is to engineer protocols or rules that seek to ensure cooperation is followed [46].

The origins of reputation systems come from behaviour in groups with humans being adept at using reputation to assess the integrity of others [34], as a means to promote their survival. This allows groups to function and humans are adept at creating heuristics, or cognitive short cuts, that allow them to find potential cooperators without extensive deliberation. However these cognitive short cuts can also have negative implications. In the context of driving dynamics for example, the type of vehicle, its manufacturer, the age, gender or other characteristics of the driver may well influence whether one driver helps another. While this may appear insignificant, in the wider human context this behaviour can have a major impact, being responsible for bias that fuels stereotyping [12], resulting in potentially unwarranted discrimination and the spread of prejudice [26]. Divisive social consequences may result [15], leading to categorisation, where the reputation that an individual incurs has no alignment to their actual behaviour. This is a key component in theories concerning intergroup conflict. These issues are also transferred to technological scenarios, depending on the capacity of machines to align with human bias or foster it themselves [43].

### 1.1 Contribution

Our contribution in this work is a new and flexible framework that allows us to explore how the sharing of reputation, by means of shared traits, affects cooperation. The approach used involves agent-based simulation, where agents have some freedom in how they adapt their behaviour based on probabilistically copying the strategy of others, based on their success. This approach allows us to explore conditions that either promote or impede cooperation. It should not be confused with agent based approaches in knowledge engineering, where protocols are sought that allow cooperation to be enforced based on individual behaviour (e.g., [46]). It can be noted that the vast majority of psychological treatments of stereotyping focus on the single trait case, despite increasing demands to capture the ground truth of social organisation [4].

Given very limited treatment of stereotyping from a complex systems perspective, the authors model it with abstraction, seeking to quantify the effects of stereotyping initially in general terms. Indirect reciprocity is the basis for our model, but other forms of cooperation could also be applied. The approach is novel because models of indirect reciprocity conventionally assume that each individual is represented by a unique reputation: in other words an individual's behaviour is entirely judged by their own actions. Stereotyping disrupts this one-to-one mapping, resulting in reputations being implicitly shared by different actors. In the context of cooperation, this means that individuals become dependent on the donation behaviour of others for an element of their reputation.

Furthermore, our framework does not assume that "groups" to which individuals belong are mutually exclusive. Reputations are calculated on traits, any number of which can be held by an individual. This better represents the fluidity that is seen in the real world, where individuals are rarely totally defined by a single group affiliation, but may be represented as a combination of characteristics and affiliations. The paper examines how both repeated sharing of the same trait, and sharing across multiple traits, affects the emergence of cooperation. This provides a mechanism to assess the cost associated with stereotyping, in terms of the effect on cooperation. To the best of our knowledge, no such previous insight has been made in this direction.

## 2 Key related literature

This research focuses on indirect reciprocity, groups and the role of reputation. Indirect reciprocity is frequently considered in the context of the donation game, where an agent has to make a decision on whether or not to provide a donation. This results in a cost $c$ to the donor, and a benefit $b$ to the recipient, and necessarily $c < b$ [5, 24]. Reputation systems act to signal an agent's overall donation behaviour to the wider population. Because other agents may use an agent's reputation in deciding when or not to donate, there is an incentive for all potential recipients to maintain reputation at a sufficient level to yield future donations [9, 21, 41].

Critical within reputation systems are assessment rules. These are the criteria by which a donor's reputation is adjusted in light of their actions, and therefore govern the extent of reward over penalty. In this sense they have been considered as a model for morality [1]. Three main alternatives for assessment of cooperative action are *image scoring*, *standing* and *judging*. The first development was standing [33], which was originally conceived for binary reputations. This assessment rule effectively classifies each individual in the population as either good or bad, penalising the good if they donate to the bad.

Image scoring [23, 41] presented the first significant alternative, where reputation is simply incremented or decremented in response to donation or defection respectively. A limitation of image scoring is that those who choose not to cooperate with defectors may be unfairly labelled as less cooperative [17, 28]. Consequently, with their roots in the work of Sugden [33], *standing* [28] and

*judging* [6] have emerged as the alternatives that capture "legitimate shirking" [10, 24, 30]. These discrimination rules have mainly been studied assuming that reputation has a binary representation [5, 27], although this was generalised for standing in [42].

The overwhelming convention is that individuals hold their own individual reputation with similarity of reputation only introduced to address uncertainty (e.g., [17, 23]). The point of deviation from this has occurred in the biological literature, specifically concerning the plausibility of group selection [44, 32]. These models assume that individuals belong to precisely one group, and it is the group entity that determines whether or not individuals propagate to future generations. This was largely dismissed by the biological literature but was revisited when the idea of multi-level selection was proposed [45], where individual and group identity coexist and may promote cooperation [25]. Reputation systems can feature in this context, allowing individuals to potentially switch between individual and group reputations [18, 34]. However this still remains controversial [29] as an explanation for biological evolution.

Psychological processes of categorisation are well seen in human behaviour, and work relating to groups and cooperation has featured consideration of both in-group bias [11, 13] and out-group prejudice [7, 43], while not necessarily invoking the use of a group reputation. These contributions reflect the disposition of individuals to differentiate, either implicitly or explicitly, based on their strong identification with self-similar individuals [16]. Stereotyping is a related extension of this, where third party individuals are categorised together through a perception of common characteristics [12]. This is well known to be a divisive phenomenon in the human world [8, 36, 39].

In the case of reputation systems, only a few contributions consider categorisation. In [2] the impact of group reputation is considered through multi-agents. Here, the concept of group reputation is shared by all individuals within a group when they interact with out-group members. This is calculated as the average of all individual reputations in a group, and assumes that group reputation is an aggregation of the behaviour of individuals. Similarly in [18], a group structure is proposed where individuals interact within their groups using a personal reputation. When they play out-group, individuals adopt a group-level reputation. This model also assumes that reputation is binary. These models do not allow for individuals to share subsets of traits, or aspects of their identity, and depend on individuals belonging to a single group. Our approach is to allow individuals to have a more complex composition of their identity, based on assessment of multiple traits against which reputations are maintained.

## 3 Model

The simulation model that is introduced pays attention to the structure of reputation that agents hold when engaged in a cooperative dilemma (indirect reciprocity). Rather than individuals necessarily holding their own unique reputation, or being identified by a single group membership, the concept of *traits* is

introduced to represent how individuals may be perceived as belonging to groups and judged through stereotypes. Traits are immutable features that are held by agents, and represent identifiable characteristics. All agents have at least one trait, and each trait may belong to one or more agents.

Rather than reputation being associated with individual agents or mutually exclusive groups, it is assumed that each trait $t$ has associated with it a reputation $r_t$, and an agent $i$ derives its personal reputation $r^i$ from the reputations of the traits associated with $i$. Specifically, for an agent $i$, let $T_i$ denote its associated set of traits, and then $r^i = \sum_{t \in T_i} r_t / |T_i|$. In other words, an agent's reputation is the average of the reputation of its associated traits.

This arrangement allows stereotyping to be considered: traits belonging to an agent and shared by others are used as a proxy for their individual reputation. Furthermore, traits do not necessarily partition agents into mutually exclusive sets or groups, providing a useful generalisation. This approach is applied using cooperation in the form of indirect reciprocity.

**Indirect reciprocity.** The donation game is adopted, which is a subclass of the mutual aid game [33] where the donor incurs a cost with no guarantee of reciprocation from the beneficiary, or any other individual. This is modelled through prosocial donations which result in a cost $c$ to the donor agent and a benefit $b$ to the recipient, where $b > c > 0$. There are wide ranging models for indirect reciprocity (e.g., [17, 23, 27, 38]), however this work uses the recent and remarkably simple approach of *social comparison* of reputation [42]. This follows the human disposition to make relative judgements about the standing of others.

Each agent $i$ carries a binary vector of variables $(s_i, u_i, d_i)$ which represents $i$'s current *action rule* with respect to $i$'s donation behaviour when it is called upon to consider making a donation to another agent $j$. The action rule indicates whether or not $i$ donates when similarity ($s_i$), upward ($u_i$), or downward self-comparison ($d_i$) is observed by $i$ in respect of $j$'s reputation ($r^j$), as compared to $i$'s own reputation value ($r^i$). Similarity in self-comparison is identified when $r^j = r^i$, upward self-comparison occurs when $r^j > r^i$, and downward self-comparison occurs when $r^j < r^i$.

Periodically each agent updates its action rule through social learning, as a consequence of observing others in the population. It is known [42] that evolution promotes the action rule $(1, 1, 0)$, allowing agents to discriminate against those having a lower reputation than themselves, thereby representing a relative threat.

**Updating reputation.** Every time an agent $i$ is called to play the donation game with a potential recipient $j$, $i$'s donation decision depends on the agent's action rule and reputation is updated as a consequence. The concept of standing is used. Specifically, it is assumed that agent $i$ has a set of traits $T_i$. If $i$ donates, then $r_t$ is incremented, for all $t \in T_i$. If $r^j \geq r^i$ and $i$ defects then the reputation of trait $t$, $r_t$ is decremented, for all $t \in T_i$. This means that an individual's actions equally affect the traits by which it is represented. Note that the updating approach ensures that a reduction in reputations does not occur when $i$ fails to

donate and $j$ is of a lesser reputation, providing a defence against shirkers. Each trait's reputation is allowed to vary in the integer range [-5,5].

**Performing the game.** The donation game is performed on a set of agents $A$ representing a population of individuals, in this case $|A| = 100$. Each agent $i$ has four key fundamental attributes: its set of traits $T_i$, its action rule $(s_i, u_i, d_i)$, its reputation $r^i$ and its fitness $f_i$. Note that $r^i = \sum_{t \in T_i} r_t / |T_i|$, and $r^i$ allows agents to play the donation game and fitness represents the accumulation of costs and benefits that are paid and received by $i$ over a generation. A generation involves making 5,000 random selections of a potential recipient $j$, from the population, to play the donation game. For an agent $j$, the potential donor agent $i$ is selected from the sub-population having at least one trait from $T_j$, with probability $s$. Here $s$ is a global parameter (not to be confused with $s_i$) that governs the extent to which an agent is disposed to playing in-group (i.e., with similar others).

At the end of a generation, reproduction occurs. This can be thought of as social learning where agents probabilistically copy the action rules of others, taking into account the success of other agents based on their fitness. Specifically, each agent $i$ in the population copies the action rule of another agent $j$ with probability $f_j / \sum_{k=1}^{n} f_k$, upon which $i$ adopts $j$'s action rule for the next generation.

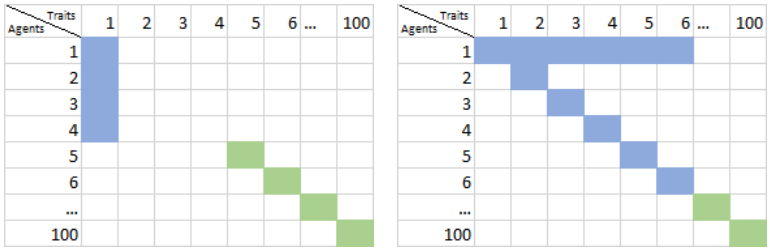
At this point mutation is applied to each element of an action rule with probability 1/100. Prior to commencing a new generation, fitness $f_i$ is set to zero ($f_i = 0, \forall i$) and for all traits $t$, $r_t = 0$ is set. Throughout a $c/b$ ratio of 0.7 is applied. 100,000 generations are performed and the simulation is principally evaluated by comparing the total number of instances of cooperation (i.e., $i$ donating to $j$ in a donation game) across all generations. Average figures of cooperation over 5 randomly seeded runs are used.

## 4 Experiments

The model provides the option for agents to experience different types of reputation sharing with other agents. An agent is *dependent* if it shares at least one trait with another agent. Otherwise the agent is *independent*. If an agent $i$ is such that $|T_i| > 1$ then $i$ is a *multi-trait* agent. Otherwise $i$ is a *single-trait* agent. Dependent agents provide the means for stereotyping to take hold. This is investigated in two ways: firstly the effect of dependent single-trait agents on the evolution of cooperation (Section 4.1); secondly the effect of a dependent multi-trait agent on the evolution of cooperation (Section 4.2).

### 4.1 Dependent single-trait agents

This section considers the effects of a single common trait $t$ being shared by a set of single-trait agents. Let $G_t$ be the set of all agents $i$ having $T_i = \{t\}$. Specifically, the maximum size of $G_t$ is determined through which cooperation can be sustained. Note that if all agents are single-trait and independent, their




(a) $|G_t|$=4 single-trait dependent agents who share trait $t = 1$.

(b) One multi-trait dependent agent, and five single trait dependent agents.

**Fig. 1.** Alternative agent-trait relationships for single-trait and multi-trait agents.

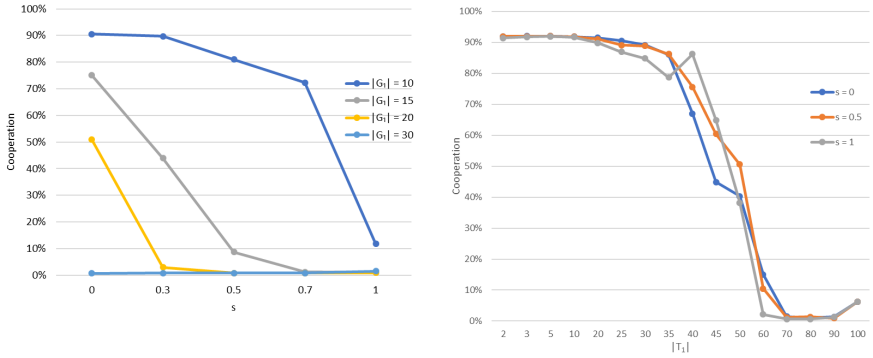
reputation is based entirely on their own past interactions and the results in [42] are replicated. At the other extreme, if all agents are dependent and share a single trait, then agents are (almost) entirely judged on the actions of others, and a greater incentive to defect is expected. The format of this experiment is visualised in Figure 1(a) and the results are shown in Figure 2(a). Two patterns emerge: firstly cooperation declines rapidly after 15 dependent single-trait agents share a common trait. Secondly, the average cooperation declines as $s$ increases.

The lack of a distinguishable personal reputation for dependent single-trait agents means that the reputational benefit of donation is shared with others, but the cost is borne by the individual. This provides an opportunity for defective strategies to take hold, where free riders can benefit from enjoying a shared reputation without donating. However this cannot be sustained at scale, leading to the global collapse of cooperation. In fact, as the number of dependent agents increases, the reputation of the shared trait can also increase in value. This leads to greater exploitation by free riders.

Figure 3, shows the action rules of agents across different values of $s$, comparing instances of the defector strategy $(0, 0, 0)$ with instances of the discriminator strategy $(1, 1, 0)$; which has been known to be dominant when all agents carry their own unique reputation [42]. Prioritising interaction with those who share the same trait (i.e., high $s$) accelerates the collapse of cooperation further as the discriminative strategy directs donations towards agents with similar reputation. When $s$ is low, dependent single-trait agents interact mainly with those who don't share their reputation as they are still incentivised to adopt cooperative strategies to maximise their fitness with a reduced risk of exploitation.

### 4.2 Dependent multi-trait agents

This section considers the effect of introducing a single dependent multi-trait agent in a population of single-trait agents, as shown in Figure 1(b). The results (Figure 2(b)) show that as the size of the set $T_1$ of the multi-trait agent




(a) Dependent single-trait agents. (b) Dependent multi-trait agent.

**Fig. 2.** Figure (a) shows the relationship between cooperation, parameter $s$, and the size of the set $G_t$ of agents sharing a common trait (see Figure 1(a)). Figure (b) shows the effect of increasing the size of the set of traits $T_1$ of a single multi-trait agent on cooperation, in a scenario where all other agents are single trait (see Figure 1(b)).

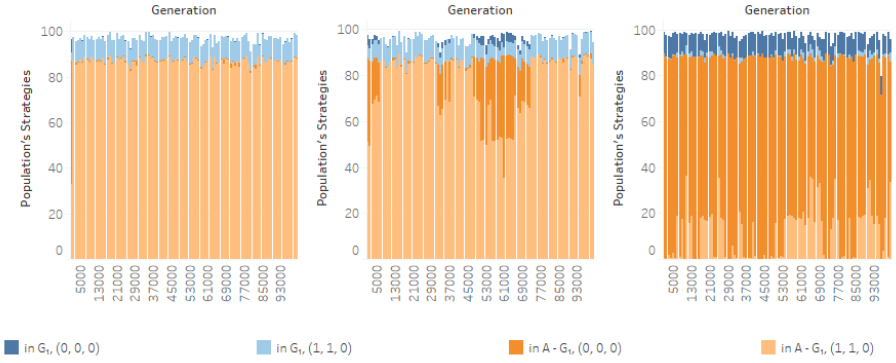



**Fig. 3.** Distribution of action rules $(0,0,0)$ and $(1,1,0)$ by generation for the sets of single-trait dependent agents $G_t$ and independent agents $A - G_t$. $|G_t| = 10$ and $s = 0$ (left), 0.5 (middle), and 1 (right).

increases, cooperation diminishes. The sharing of the multi-trait agent's reputation is dispersed across single-trait agents who between themselves have no trait in common. This helps to suppress the rise of defective action rules, as compared to the previous scenario (Section 4.1). In fact, $|T_1|$ can reach a considerable size (e.g., 30-35 traits) before which cooperation starts to significantly diminish.

In this scenario, single-trait dependent agents rely entirely on themselves and the multi-trait agent for their reputation. Each single-trait dependent agent can also free ride on the single multi-trait agent, and this opens the opportunity for defection to establish itself, although to a lesser extent than the case presented

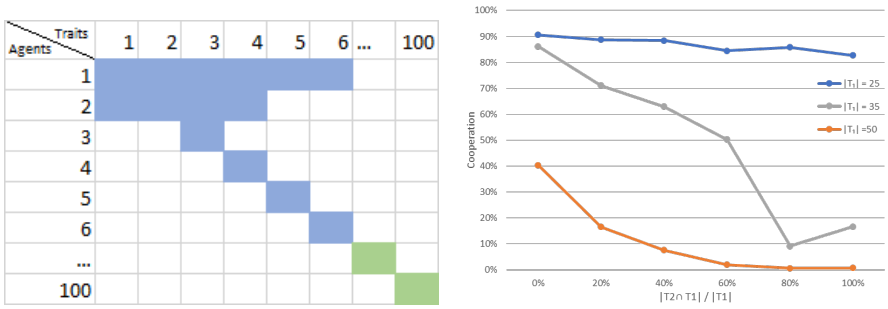



**Fig. 4.** The figures show the relationship between agents and traits for two dependent multi-trait agents (left) and the average cooperation produced as a function of the size of the intersection between the sets belonging to agents one and two for different values of $|T_1|$ where $s = 0$ (right).

in Section 4.1. When the number of traits of the multi-trait agent is relatively small, the presence of free riding dependent single-trait agents can be sustained without too much disruption to the reputation of the multi-trait agent. As $|T_1|$ increases, and the number of dependent single-trait agents increases, there is a greater opportunity for free-riding action rules to take hold. At the same time, there are fewer independent single trait agents available in the population. This promotes the collapse of cooperation. As soon as a defective strategy takes hold across the population, it then opens the opportunity for this to spread to other agents. Interestingly, s has relatively little impact on whether dependent agents prioritise playing with those that have a common trait. However, they are less likely to have equal reputation in this instance.

Finally a second multi-trait agent is added, by replacing a single-trait agent (agent number 2) in Figure 1(b), where $T_2 \subseteq T_1$. Figure 4 shows the effect of varying $|T_2 \cap T_1|$, that is the extent to which $T_2$ has the same traits as $T_1$. These results show that high proportions of shared identity through multi-trait agents undermine the reputation system. Because the second multi-trait agent can hold a large subset of the first agent's traits, it can heavily disrupt the first agent's reputation, by using defection as its action rule. This effect is more pronounced than that of a dependent single trait agent sharing reputation with the multi-trait dependent agent, and increases as $|T_2 \cap T_1|$ increases.

## 5 Discussion and Conclusion

Through a general framework for considering reputation, the authors have found that reputation systems for cooperation are heavily disrupted by the sharing of reputation through common traits. Stereotyping takes place, where traits are used as proxy for indirectly assessing an individual's reputation. This introduces the opportunity for agents to disconnect their actions from their reputation. Agents can deploy defective strategies: that is an agent can avoid paying the full

costs of donation but receives donations based on the reputation aligning with its associated traits. How the reputation is shared, through inheritance of traits, is highly influential. Holding multiple traits presents an opportunity for agents to share a limited proportion of their identity with others. In doing so they have the potential to better control their exposure to defectors.

Single-trait and multi-trait agents are differentiated in how other agents can share their traits. Under uniform conditions, single-trait agents have a reduced chance of others having a trait in common. However, when another agent shares their trait, their reputation becomes susceptible to the actions of a third party. In contrast, for multi-trait agents, increasing the number of traits can give them a chance to retain an element of unique personal identity, through traits that aren't shared with others. Moreover, for multi-trait agents, sharing can occur with a number of agents that have no dependency between them, in terms of common traits.

Given the enormous number of possible ways in which traits can be shared, in this work our focus has concerned assessing basic aspects of sharing, surrounding the number of traits held by an agent. The results show that reasonable levels of cooperation can be sustained while there is a modest level of sharing of identity in the population, after which cooperation collapses.

This highlights the importance of individual versus group identity in reputation systems. It also warrants further investigation, being an important issue relevant to identity fusion [35], where effectively an individual's personal identity becomes identical to that of the group - in other words, distinguishable personal traits diminish. Understanding the extent to which personal identity, group identity and cooperation trade-off against each other, is an important future goal. The exploratory work in this paper validates the framework presented as a means to accomplish this.

**Acknowledgements.** W. Bedewi is funded by King Abdulaziz University, Saudi Arabia. Additionally the research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes not with standing any copyright notation hereon. This research was also supported by the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via Welsh Government.

## References

1. Alexander, R.D.: The biology of moral systems. Transaction Publishers (1987)
2. Baranski B. et al.: The impact of group reputation in multiagent environments. In: Proc. of CEC '06. pp. 1224–1231 (2006)

3. Bear, A., Rand, D.G.: Intuition, deliberation, and the evolution of cooperation. Proceedings of the National Academy of Sciences **113**(4), 936–941 (2016)
4. Bowleg, L.: Intersectionality: an underutilized but essential theoretical framework for social psychology. In: The Palgrave handbook of critical social psychology, pp. 507–529. Springer (2017)
5. Brandt, H., Ohtsuki, H., Iwasa, Y., Sigmund, K.: A survey of indirect reciprocity. In: Mathematics for Ecology and Environmental Sciences, pp. 21–49. Springer (2007)
6. Brandt, H., Sigmund, K.: The logic of reprobation: assessment and action rules for indirect reciprocation. J. of Theoretical Biology **231**(4), 475–486 (2004)
7. Brewer, M.B.: The psychology of prejudice: Ingroup love and outgroup hate? J. of social issues **55**(3), 429–444 (1999)
8. Dovidio, J.F., Gaertner, S.L., Validzic, A.: Intergroup bias: status, differentiation, and a common in-group identity. J. of personality and social psychology **75**(1), 109 (1998)
9. Fehr, E.: Human behaviour: don't lose your reputation. Nature **432**(7016), 449–450 (2004)
10. Fishman, M.A.: Indirect reciprocity among imperfect individuals. J. of Theoretical Biology **225**(3), 285–292 (2003)
11. Fu, F., Tarnita, C.E., Christakis, N.A., Wang, L., Rand, D.G., Nowak, M.A.: Evolution of in-group favoritism. Scientific reports **2**, 460 (2012)
12. Galinsky, A.D., Moskowitz, G.B.: Perspective-taking: decreasing stereotype expression, stereotype accessibility, and in-group favoritism. J. of personality and social psychology **78**(4), 708 (2000)
13. Hammond, R.A., Axelrod, R.: The evolution of ethnocentrism. J. of Conflict Resolution **50**(6), 926–936 (2006)
14. Imbsweiler, J., Ruesch, M., Weinreuter, H., León, F.P., Deml, B.: Cooperation behaviour of road users in t-intersections during deadlock situations. Transportation research part F: traffic psychology and behaviour **58**, 665–677 (2018)
15. Kawakami, K., Amodio, D.M., Hugenberg, K.: Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In: Advances in experimental social psychology, vol. 55, pp. 1–80. Elsevier (2017)
16. Launay, J., Dunbar, R.I.M.: Playing with strangers: Which shared traits attract us most to new people? PLOS ONE **10**, 1–17 (06 2015)
17. Leimar, O., Hammerstein, P.: Evolution of cooperation through indirect reciprocity. Proc. of the Royal Society B: Biological Sciences **268**(1468), 745–753 (2001)
18. Masuda, N.: Ingroup favoritism and intergroup cooperation under indirect reciprocity based on group reputation. J. of Theoretical Biology **311**, 8–18 (2012)
19. Melnik, M.I., Alm, J.: Does a seller's ecommerce reputation matter? evidence from ebay auctions. The J. of industrial economics **50**(3), 337–349 (2002)
20. de Melo, C.M., Marsella, S., Gratch, J.: Human cooperation when acting through autonomous machines. Proceedings of the National Academy of Sciences **116**(9), 3482–3487 (2019)
21. Milinski, M., Semmann, D., Krambeck, H.J.: Reputation helps solve the 'tragedy of the commons'. Nature **415**(6870), 424–426 (2002)
22. Molleman, L., van den Broek, E., Egas, M.: Personal experience and reputation interact in human decisions to help reciprocally. Proc. R. Soc. B **280**(1757), 20123044 (2013)
23. Nowak, M.A., Sigmund, K.: Evolution of indirect reciprocity by image scoring. Nature **393**(6685), 573–577 (1998)

24. Nowak, M.A., Sigmund, K.: Evolution of indirect reciprocity. Nature **437**(7063), 1291–1298 (2005)
25. Nowak, M.A., Tarnita, C.E., Wilson, E.O.: The evolution of eusociality. Nature **466**(7310), 1057 (2010)
26. Oakes, P.J., Turner, J.C.: Social categorization and intergroup behaviour: Does minimal intergroup discrimination make social identity more positive? European J. of Social Psychology **10**(3), 295–301 (1980)
27. Ohtsuki, H., Iwasa, Y.: The leading eight: social norms that can maintain cooperation by indirect reciprocity. J. of Theoretical Biology **239**(4), 435–444 (2006)
28. Panchanathan, K., Boyd, R.: A tale of two defectors: the importance of standing for evolution of indirect reciprocity. J. of Theoretical Biology **224**(1), 115–126 (2003)
29. Pinker, S.: The false allure of group selection (2012)
30. Rand, D.G., Nowak, M.A.: Human cooperation. Trends in Cognitive Sciences **17**(8), 413–425 (2013)
31. Resnick, P., Kuwabara, K., Zeckhauser, R., Friedman, E.: Reputation systems. Communications of the ACM **43**(12), 45–48 (2000)
32. Smith, J.M.: Group selection and kin selection. Nature **201**(4924), 1145 (1964)
33. Sugden, R.: The economics of rights, co-operation and welfare. Blackwell Oxford (1986)
34. Suzuki, S., Akiyama, E.: Reputation and the evolution of cooperation in sizable groups. Proc. of the Royal Society B: Biological Sciences **272**(1570), 1373–1377 (2005)
35. Swann Jr, W.B., Gómez, Á., Seyle, D.C., Morales, J., Huici, C.: Identity fusion: the interplay of personal and social identities in extreme group behavior. J. of personality and social psychology **96**(5), 995 (2009)
36. Tajfel, H., Billig, M.G., Bundy, R.P., Flament, C.: Social categorization and intergroup behaviour. European J. of social psychology **1**(2), 149–178 (1971)
37. Tajfel, H., Turner, J.C.: An integrative theory of intergroup conflict. The social psychology of intergroup relations **33**(47), 74 (1979)
38. Takahashi, N., Mashima, R.: The importance of subjectivity in perceptual errors on the emergence of indirect reciprocity. J. of Theoretical Biology **243**(3), 418–436 (2006)
39. Turner, J.C., Hogg, M.A., Oakes, P.J., Reicher, S.D., Wetherell, M.S.: Rediscovering the social group: A self-categorization theory. Basil Blackwell (1987)
40. Wasko, M.M., Faraj, S.: "it is what one does": why people participate and help others in electronic communities of practice. The J. of Strategic Information Systems **9**(2-3), 155–173 (2000)
41. Wedekind, C., Milinski, M.: Cooperation through image scoring in humans. Science **288**(5467), 850–852 (2000)
42. Whitaker, R.M., Colombo, G.B., Allen, S.M., Dunbar, R.I.: A dominant social comparison heuristic unites alternative mechanisms for the evolution of indirect reciprocity. Scientific reports **6**, 31459 (2016)
43. Whitaker, R.M., Colombo, G.B., Rand, D.G.: Indirect reciprocity and the evolution of prejudicial groups. Scientific reports **8**(1), 13247 (2018)
44. Wilson, D.S.: A theory of group selection. Proc. of the national academy of sciences **72**(1), 143–146 (1975)
45. Wilson, D.S., Sober, E.: Reintroducing group selection to the human behavioral sciences. Behavioral and brain sciences **17**(4), 585–608 (1994)
46. Wu, J., Balliet, D., Van Lange, P.A.: Reputation, gossip, and human cooperation. Social and Personality Psychology Compass **10**(6), 350–364 (2016)