# Dynamics of Fairness in Groups of Autonomous Learning Agents

Fernando P. Santos[1,2(✉)], Francisco C. Santos[1,2], Francisco S. Melo[1], Ana Paiva[1], and Jorge M. Pacheco[2,3]

[1] INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Taguspark, Av. Prof. Cavaco Silva, 2780-990 Porto Salvo, Portugal
fernando.pedro@tecnico.ulisboa.pt
[2] ATP-Group, 2780-990 Porto Salvo, Portugal
[3] CBMA and Departamento de Matemática e Aplicações, Universidade do Minho, Campus de Gualtar, 4710-057 Braga, Portugal

**Abstract.** Fairness plays a determinant role in human decisions and definitely shapes social preferences. This is evident when groups of individuals need to divide a given resource. Notwithstanding, computational models seeking to capture the origins and effects of human fairness often assume the simpler case of two person interactions. Here we study a multiplayer extension of the well-known Ultimatum Game. This game allows us to study fair behaviors in a group setting: a proposal is made to a group of Responders and the overall acceptance depends on reaching a minimum number of individual acceptances. In order to capture the effects of different group environments on the human propensity to be fair, we model a population of learning agents interacting through the multiplayer ultimatum game. We show that, contrarily to what would happen with fully rational agents, learning agents coordinate their behavior into different strategies, depending on factors such as the minimum number of accepting Responders (to achieve group acceptance) or the group size. Overall, our simulations show that stringent group criteria leverage fairer proposals. We find these conclusions robust to (*i*) asynchronous and synchronous strategy updates, (*ii*) initially biased agents, (*iii*) different group payoff division paradigms and (*iv*) a wide range of error and forgetting rates.

## 1 Introduction

Fairness plays a central role in human decision-making and it often directs the actions of people towards unexpected outcomes. This fact has puzzled academics from multiple fields and the subject comprises a fertile ground of multidisciplinary research [9,10]. A neat way to verify that humans often give up their own material gains in order to achieve fair outcomes is achieved by observing how people play a very simple game named the Ultimatum Game (UG) [13]. In this game, two players interact with each other. The Proposer is endowed with some resource and has to propose a division with the Responder. After that, the Responder has to state her acceptance or rejection. If the proposal is rejected,


N. Osman and C. Sierra (Eds.): AAMAS 2016 Ws Best Papers, LNAI 10002, pp. 107–126, 2016.
DOI: 10.1007/978-3-319-46882-2_7

none of the players earn anything. If the proposal is accepted, they will divide the resource as it was proposed. In the context of UG, the outcome of any accepted proposal stands as a social optimum (in the Pareto sense, i.e., no other player can improve her payoff without damaging the payoff of others) however, only the egalitarian division, in which both the Proposer and the Responder earn a similar reward, is considered a fair result.

A first approach, attempting to predict the behavior of people in this game, relies on the assumption that each player is a rational agent that seeks to unconditionally maximize the rewards. In this case, it is easy to notice that the Responder should always accept any offer; wherefore, the Proposer should never fear to have a proposal rejected and should always propose the minimum possible. This indeed constitutes the sub-game perfect equilibrium of the UG [27]. A vast number of works, however, report experiments with people in which the rational sub-game prediction is not played and fair outcomes are verified [13,26,35,46]. Humans tend to reject low proposals and manage to offer high/fair divisions. Offers are higher than expected even in the so-called Dictator games, where the Responders do not have the opportunity to reject and proposals are always accepted [10,16].

If one intends to model, explain and tentatively predict the behavior of people in this game, new mathematical and computational tools have to be employed, other than the game theoretical sub-game perfect equilibrium. For instance, by relaxing the rational assumption made about human decision making and by simply undertaking that strategies are adopted or renounced by individual [11,22,43] or social learning [31,33,41], various mechanisms can be tested and different conclusions can be obtained. In the second case, assuming that agents co-habit a population and adopt strategies with a probability that grows with the success those strategies are perceived to consent, the dynamics of strategy adoption interestingly resemble a process of gene replication and the evolving behavior of agents can be modeled by tools from Evolutionary Game Theory (EGT) [50]. Interestingly, EGT can also be used to study individual learning dynamics [1].

UG has been studied in the context of EGT and it has been shown that when Proposers collect pieces of information about opponents' previous actions, it is worth for the Responders to cultivate a fierce bargainer reputation [25]. This explains the long-term benefits of Responders that acquire an intransigent image by rejecting unfair offers. Other models attribute the evolution of fairness to repeated interactions [49], to empathy [29] or even simply to environmental noise and stochasticity [32,37]. A slightly different approach suggests that fair Proposers and Responders may emerge due to the topological arrangement of their network of contacts: if individuals are arranged in lattices [30,45] or complex networks [17,42] clusters of fairness may emerge.

While the UG is ubiquitous in real-life encounters, there is a wide range of human interactions that a pairwise interaction model does not enclose. It is perfectly straightforward to realize that also UG instances take place in groups, with proposals being made to assemblies [38]. Take the case of pervasive democratic

institutions, economic and climate summits, markets, auctions, or the ancestral activities of proposing divisions regarding the loot of group hunts and fisheries. All those examples go beyond a pairwise interaction. More specifically, the relation of groups and the possibility of fair allocations is a topic utterly relevant in the context of group buying [19,20], collective bargaining of work contracts or coalition policy making [14] and indeed, there is a growing interest in doing experiments with multiplayer versions of fairness games [3,6,7,9,12,23,39]. A simple extension of UG may turn it adequate to study a wide variety of ubiquitous group encounters. This extension, the Multiplayer UG (MUG), allows to study the traditional UG in a context where proposals are made to groups that should decide about its acceptance or rejection [38]. In the context of this game, some questions need to be addressed: What is the role of the specific group environment on people's behavior? What is the impact of group acceptance rules (i.e., the minimum number of accepting Responders to obtain group acceptance) on individual offers? What is the role of group size on fairness?

If one assumes that agents always opt for the payoff maximizing strategy, the previous questions have trivial answers: proposals in MUG are accepted regardless of the group particularities and any effect of group size or group acceptance rules in the preferred people's behavior should be neglected. However, abandoning this strong rationality assumption, and acknowledging that unexpected behaviors often result from an adaptive process of evolution and learning, turns plausible that different group environments can shape decision making and nurture fair outcomes. As mentioned before, the use of multiagent learning techniques can, for that end, unveil important characteristics of human interactions that are neglected by typical equilibrium analysis. Here we seek to analyze the role of different group environments on the emergence of fair outcomes, by combining MUG with agents that adapt their behavior through reinforcement learning [44]. We implement and test with the well-known Roth-Erev reinforcement learning algorithm [35]. We show that group size and different group acceptance rules impact, in a nontrivially manner, the learned strategies and the associated fairness: increasing the minimum number of accepting Responders to achieve group acceptance has the effect of increasing the offered values and consequently fairness; secondly, the effect of group size strongly depends on the group acceptation threshold.

For simplicity and readability purposes, in Table 1 we provide a list of the nomenclature used through this document. In Sect. 2, we present the MUG [38] and we review the equilibrium notions of classical game theory, namely, the subgame perfection. In Sect. 3 we present the Roth-Erev learning model that we use thorough this work. After that, in Sect. 4, we present the results showing that, within a population of adaptive agents, group environment (group acceptance rules and group size) indeed plays a fundamental role in the employed strategies. In Sect. 5 we discuss the obtained results and provide a set of concluding remarks.

**Table 1.** Glossary

| Symbol | Meaning |
|---|---|
| $p$ | Offer by Proposer |
| $q$ | Acceptance threshold of Responder |
| $\Pi_P(p_i, q_{-i})$ | Payoff earned by a Proposer |
| $\Pi_R(p_j, q_{-j})$ | Payoff earned by a Responder |
| $\Pi(p_i, q_i, p_{-i}, q_{-i})$ | Payoff being Proposer and Responder |
| $a_{p_i,q_{-i}}$ | Group acceptance flag |
| $Q(t)$ | Propensity matrix at time $t$ |
| $\lambda$ | Forgetting rate |
| $\epsilon$ | Local experimentation |
| $\rho_{ki}(t)$ | Probability that $k$ uses strategy $i$ |
| $\bar{p}$, $\bar{q}$ | Average $p, q$ population-wide |
| $i_{p,q}$ | Integer representation of strategy $(p, q)$ |
| $R$ | Number of runs |
| $Z$ | Population size |
| $N$ | Group size |
| $M$ | Group acceptance rule |
| $T$ | Number of time steps |
| $R$ | Number of runs |

## 2 Multiplayer Ultimatum Game

Often people incur in interactions that are fundamentally rooted in proposals made to groups. These proposals can naturally be accepted or rejected, depending on the subsequent bargaining and group acceptance rules. The outcome of this interaction can favor unequally each part involved and is thereby likely that concerns about fairness puzzle each player mood. The role played by the group in this interaction is overlooked by the traditional two-person UG. Thereby, here we present and analyze the Multiplayer Ultimatum Game (MUG) which allows us to test the effect of different group environments on the behaviors adopted by people and the associated fairness levels [38].

In the UG, we can assume that the strategy of the Proposer is the fraction of resource offered to the Responder ($p$) and the strategy of the Responder is the personal threshold ($q$) used to decide about acceptance or rejection [25,30]. Only whenever $p \geq q$ the proposal is accepted. Considering that the amount being divided sums to 1, an accepted proposal of $p$ endows the Proposer with $1-p$ and the Responder with $p$. If the proposal is rejected, none of the individuals earn anything. The UG can now be extended to a N-person game if we account

for the existence of a group composed by $N-1$ Responders [36,38]. The group decision making can be arbitrarily complex yet, we simplify this process by assuming that each of the $N-1$ Responders accepts or rejects the proposal and the overall group acceptance depends on a group acceptance rule: if the number of acceptances equals or exceeds a minimum number of accepting Responders, $M$, the proposal is accepted by the group. In this case, the Proposer keeps what she did not offer $(1-p)$ and the offer is divided by the Responders (in two possible ways, as detailed next); otherwise, if the number of acceptances remains below $M$, the proposal is rejected by the group and no one earns anything. The accepted proposal can be $(i)$ evenly divided by all the Responders or $(ii)$ only divided by the accepting Responders.

The payoff function describing the gains of a Proposer $i$, with strategy $p_i$, facing a group of Responders with strategies $q_{-i} = \{q_1, ..., q_j, ..., q_{N-1}\}, j \neq i$ reads as

$$\Pi_P(p_i, q_{-i}) = (1 - p_i)a_{p_i,q_{-i}} \tag{1}$$

Where $a_{p_i,q_{-i}}$ summarizes the group acceptance of the proposal made by agent $i$ $(p_i)$, standing as

$$a_{p_i,q_{-i}} = \begin{cases} 1, & \text{if } \sum_{q_j \in q_{-i}} \Theta(p_i - q_j) \geq M. \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

$\Theta(x)$ is the Heaviside unit step function, having value 1 whenever $x \geq 0$ and 0 otherwise. This way, $\Theta(p_i - q_j) = 1$ if agent $j$ accepts agent's $i$ proposal and $\sum_{q_j \in q_{-i}} \Theta(p_i - q_j)$ is the number of Responders (within those using strategies $q_{-i} = \{q_1, ..., q_j, ..., q_{N-1}\}, j \neq i$) accepting proposal $p_i$.

Similarly, the payoff function describing the gains of a Responder belonging to a group with a strategy profile $q_{-j} = \{q_1, ..., q_k, q_i, ..., q_{N-1}\}, k \neq j$, listening to a Proposer $j$ with strategy $p_j$, is, in the case of proposals evenly divided by all, given by

$$\Pi_R(p_j, q_{-j}) = \frac{p_j}{N-1} a_{p_j,q_{-j}}. \tag{3}$$

In the case of proposals divided by the accepting Responders, the payoff of a Responder (with strategy $q_i$) is given by

$$\Pi_R(p_j, q_{-j}) = \frac{p_j \Theta(p_j - q_i) a_{p_j,q_{-j}}}{\sum_{q_k \in q_{-j}} \Theta(p_j - q_k)} \tag{4}$$

This equation implies that a Responder only earns something if both she and the group accept that proposal. In turn, Eq. (3) implies that only the group has to accept a proposal, for any Responder to earn something.

We can assume that these games occur in groups where each individual acts once as Proposer and $N-1$ times as Responder. This way, the overall payoff of an individual with strategy $(p_i, q_i)$, playing in a group with strategy profile $(p_{-i}, q_{-i})$, is given by

$$\Pi(p_i, q_i, p_{-i}, q_{-i}) = \Pi_P(p_i, q_{-i}) + \sum_{p_j \in p_{-i}} \Pi_R(p_j, q_{-i}) \tag{5}$$

The interesting values of M range between 1 and $N-1$. If $M = 0$, no Responders are needed to accept a proposal and so, all proposals would be accepted. With $M > N - 1$ all proposals are rejected irrespectively of the strategies used by the players.

### 2.1 Sub-game Perfect Equilibrium

In order to derive the sub-game perfect equilibrium of MUG, let us introduce some canonical notation. A game given in a sequential form has a set of stages in which a specific player (chosen by a *player function*) should act. A *history* stands as any possible sequence of actions, given the turns assigned by the player function. Roughly speaking, a *terminal history* is a sequence of actions that go from the beginning of the game until an end, after which there are no actions to follow. Each *terminal history* will prescribe different outcomes to the players involved, given a specific *payoff* structure that fully translates the preferences of the individuals. This way, a *sub-game* is composed by the set of all possible histories that may follow a given non-terminal history. A strategy profile is a *sub-game perfect equilibrium* if it also the Nash equilibrium of every sub-game [27].

Let us turn to the specific example of MUG to clarify this idea. In this game, the Proposer does the first move and the Responders should, secondly, state acceptance or rejection. The game has two stages and any terminal history is composed by sets of two actions, one taken by a single individual (Proposer, with possibility to suggest any division of the resource) and the second by the group (acceptance or rejection).

Picture the scenario in which groups consist of 5 players, where one is the Proposer, the other 4 are the Responders and M = 4 (different $M$ would lead to the same conclusions). Let us evaluate two possible strategy profiles: $s_1 = (0.8, 0.8, 0.8, 0.8, 0.8)$ and $s_2 = (\mu, 0, 0, 0, 0)$, where the first value is the offer by the Proposer and the remaining 4 are the acceptance thresholds by the Responders. Both strategy profiles are Nash Equilibria of the whole game. In the first case, the Proposer does not have interest in deviating from 0.8: if she lowers this value, the proposal will be rejected and thus she will earn 0; if she increases the offer, she will keep less to herself. The same happens with the Responders: if they increase the threshold, they will earn 0 instead of 0.2, and if they decrease it, nothing happens (non-strict equilibrium). The exact same reasoning can be made for $s_2$, assuming that $\mu/(N-1)$ is the smallest possible division of the resource.

Regarding sub-game perfection, the conclusions are different. Assume the *history* in which the Proposer has chosen to offer $\mu$ (let us call the sub-game after this history, in which only one move is needed to end the game, $h$). In this case, the payoff yielded by $s_1$ is $(0, 0, 0, 0, 0)$ (every Responder rejects a proposal of $\mu$) and the payoff yielded by $s_2$ is $(1-\mu, \mu/(N-1), \mu/(N-1), \mu/(N-1), \mu/(N-1))$. So it pays for the Responders to choose $s_2$ instead of $s_1$, which means that $s_1$ is not a Nash Equilibrium of the sub-game $h$. Indeed, while any strategy profile in the form $s = (p, p, p, p, p), \mu < p \leq 1$ is a Nash Equilibrium of MUG, only $s^* = (\mu, 0, 0, 0, 0)$ is the sub-game perfect equilibrium. As described in the

introductory section, a similar conclusion, yet simpler and more intuitive, could be reached through backward induction.

## 3 Learning Model

The use of multiagent learning algorithms can unveil fundamental properties of human interactions that are overlooked if one would assume that individuals always behave following fully rational behaviors [8,11,22,35,43]. Particularly, the Roth-Erev algorithm was used with remarkable success in modeling the process of human learning when playing well-known interaction paradigms such as the Ultimatum Game [35]. We use the Roth-Erev algorithm to analyze the outcome of a population with learning agents playing MUG in groups of size $N$. In this algorithm, at each time-step $t$, each agent $k$ is defined by a propensity vector $Q_k(t)$. Over time, this vector is updated given the payoff gathered after each play. Successfully employed actions will grant larger payoffs that, when added to the corresponding propensity value, will increase the probability of repeating that strategy in the future (as will be made clear below). We consider that games take place within a population with $Z$ ($Z > N$) adaptive agents. Agents earn payoff following an anonymous random matching model [11], i.e., we sample random groups without any kind of preferential arrangement or reciprocation mechanism. We consider MUG with discretized strategies. We round the possible values of $p$ (proposed offers) and $q$ (individual threshold of acceptance) to the closest multiple of $1/D$, where $D$ measures the granularity of the strategy space considered. We map each pair of decimal values $p$ and $q$ into an integer representation, thereafter $i_{p,q}$ is the integer representation of strategy $(p, q)$ and $p_i$ (or $q_i$) designates the $p$ ($q$) value corresponding to the strategy with integer representation $i$.

The core of the learning algorithm takes place in the update of the propensity vector of each agent, $Q(t+1)$, after a play at time-step $t$. Denoting the set of possible actions by $A, a_i \in A : a_i = \{p_i, q_i\}$, and the population size by $Z$, the propensity matrix, $Q(t) \in R_+^{Z \times |A|}$, is updated following the base rule

$$Q_{ki}(t+1) = \begin{cases} Q_{ki}(t) + \Pi(p_i, q_i, p_{-i}, q_{-i}) & \text{if } k \text{ played } i \\ Q_{ki}(t) & \text{otherwise} \end{cases} \tag{6}$$

The above update can be enriched with human learning features: *forgetting rate* ($\lambda, 0 \leq \lambda \leq 1$) and *local experimentation error*, ($\epsilon, 0 \leq \epsilon \leq 1$) [35], leading to an update rule slightly improved,

$$Q_{ki}(t+1) = \begin{cases} Q_{ki}(t)\bar{\lambda} + \Pi(p_i, q_i, p_{-i}, q_{-i})(1-\epsilon) & k \text{ played } i \\ Q_{ki}(t)\bar{\lambda} + \Pi(p_i, q_i, p_{-i}, q_{-i})\frac{\epsilon}{4} & k \text{ pl. } i_p \pm 1 \\ Q_{ki}(t)\bar{\lambda} + \Pi(p_i, q_i, p_{-i}, q_{-i})\frac{\epsilon}{4} & k \text{ pl. } i_q \pm 1 \\ Q_{ki}(t)\bar{\lambda} & \text{otherwise} \end{cases} \tag{7}$$

where $\bar{\lambda} = 1 - \lambda$ and $i_p \pm 1$ ($i_q \pm 1$) corresponds to the index of the $p$ ($q$) values of the strategies adjacent to $p_i$ ($q_i$), naturally depending on the discretization

chosen. The introduction of local experimentation errors is convenient as they prevent the probability of playing the less used strategies (however close to the used ones) from going to 0. Moreover, those errors may introduce the spontaneous trial of novel strategies, a feature that is both human-like and showed to improve the performance of autonomous agents [40]. The forgetting rate is convenient to inhibit the entries of $Q$ from growing without bound: when the propensities reach a certain value, the magnitude of the values forgotten, $Q_{ki}(t)\lambda$, approach those of the payoffs being added, $\Pi(p_i, q_i, p_{-i}, q_{-i})$.

When an agent is called to pick an action, she will do so following the probability distribution dictated by the normalization of her propensity vector. The probability that individual $k$ picks the strategy $i$ at time $t$ is given by

$$\rho_{ki}(t) = \frac{Q_{ki}(t)}{\sum_n Q_{ni}(t)} \tag{8}$$

The initial values of propensity, $Q(0)$, have a special role in the convergence to a given propensity vector and on the exploration *versus* exploitation dilemma. If the norm of propensity vectors in $Q(0)$ is high, the initial payoffs obtained will have a low impact on the probability distribution. Oppositely, if the norm of propensity vectors in $Q(0)$ is small, the initial payoffs will have a big impact on the probability of choosing the corresponding strategy again. Convergence will be faster if the values in $Q(0)$ are low, yet in this case agents will not initially explore a wide variety of strategies. Here we consider three variants of initially attributed values to $Q(0)$: $(i)$ random initial propensity, where each entry $Q_{ki}(0)$ assumes real values randomly sampled (uniformly) from the interval $[0, Q(0)_{max}]$; $(ii)$ propensity values initially high on the strategy $p = q = 0$ – specifically, we attribute a random value between 0 and 1 to the propensities corresponding to the strategies $p \neq 0$, $q \neq 0$ and we attribute the value $Q(0)_{max}$ to the propensity corresponding to the strategy $p = q = 0$; $(iii)$ values of propensity initially high on the strategy $p = q = 1$ and low on strategies $p \neq 1$, $q \neq 1$.

All together, the individual learning algorithm can be intuitively perceived: when individual $k$ uses strategy $i$ she will reinforce the use of that strategy provided the gains that she obtained; higher gains will increase to a higher extent the probability of using that strategy in the future. The past use of the remaining strategies, and the obtained feedbacks, will be forgotten over time; similar strategies to the one employed (which in the case of MUG are just the adjacent values of proposal and acceptance threshold) will also be reinforced, yet to a lower extent. This learning algorithm is rather popular, providing a canonical method of reinforcement learning that was successfully applied in the past to fit the way that people learn to play social dilemmas [8,35]. It is noteworthy that two important properties of human learning are captured by this model: the Law of Effect [47] and the Power Law of Practice [24]. The first poses that humans (and animals) tend to reinforce the use of previously successfully employed strategies; the Power Law of Practice states that the learning curve of a given task by a human is initially steep and, over time, gets flat. Indeed, by using the Roth-Erev algorithm, larger payoffs reinforce to a larger extent

**Algorithm 1.** Roth-Erev reinforcement learning algorithm in an adaptive population and considering **synchronous** update of propensities.

$Q(0) \leftarrow$*initialization*;
**for** $t \leftarrow 1$ **to** $T$, *total number of time-steps* **do**
  $tmp \leftarrow \{0, ..., 0\}$ `/* keeps the temporary payoffs of the current time step to allow for synchronous update of propensities */`;
  **for** $k \leftarrow 1$ **to** $Z$ **do**
    1. *pick random group with individual* $k$;
    2. *collect strategies (Eq. 8)*;
    3. *calculate payoff of* $k$ *(Eq. 5)*;
    4. *update* $tmp[k]$ *with payoff obtained*;
  *update* $Q(t)$ *given* $Q(t-1)$ *and* $tmp$ *(Eq. 7)*;
  *save* $\bar{p}$ *(Eq. 9)*;
  *save* $\bar{q}$ *(Eq. 9)*;

the usage of a given strategy (alongside preventing the usage of the remaining strategies) and, following Eq. (8), payoffs have a larger relative impact on the probability of picking a strategy at the beginning of the learning process, when $Q_{ki}(t)$ values are lower.

**Algorithm 2.** Roth-Erev reinforcement learning algorithm in an adaptive population and considering **asynchronous** update of propensities.

$Q(0) \leftarrow$*initialization*;
**for** $t \leftarrow 1$ **to** $T$, *total number of time-steps* **do**
  **for** $i \leftarrow 1$ **to** $Z$ **do**
    1. *pick random individual* $k$ *and random group with* $k$;
    2. *collect strategies (Eq. 8)*;
    3. *calculate payoff of* $k$ *(Eq. 5)*;
    4. *update* $Q_k$ *with the payoff obtained*;
  *save* $\bar{p}$ *(Eq. 9)*;
  *save* $\bar{q}$ *(Eq. 9)*;

The remaining algorithm is summarized in Algorithms 1 and 2. In Algorithm 1 we detail the synchronous version of the algorithm. In this case, we guarantee that during each time step every agent has the opportunity to update her propensity values. Moreover, during a given time step, all agents play and the obtained payoff is kept in a temporary registry, so that all agents update their propensities at once, after a time step elapses.

In Algorithm 2 we summarize the asynchronous version of the propensity updates. In this case, $Z$ (the population size) agents are randomly selected to update their propensity values during one time step, without any guarantee that all agents are given this opportunity and that no agents are repeatedly selected. Additionally, when an agent plays, the corresponding propensity value is immediately updated, precluding any kind of synchronism in the propensity update process.

We keep track of the average values of $p$ and $q$ in the population, designating them by $\bar{p}$ and $\bar{q}$. Provided a propensity matrix, they are calculated as

$$\bar{p} = \frac{1}{Z} \sum_{1<k<Z} \sum_{1<i<|A|} \rho_{ki} p_i$$

$$\bar{q} = \frac{1}{Z} \sum_{1<k<Z} \sum_{1<i<|A|} \rho_{ki} q_i \tag{9}$$

In the next section, we present and discuss the results stemming from our experiments.

## 4 Results

Through the simulation of the multiagent system described in the previous section, we first show that different group acceptance rules have a considerable impact on the average values of offers ($p$) and acceptance thresholds ($q$) learned by the population. As the time-series in Fig. 1 (left column) show, when MUG takes place in groups of size $N = 8$ and for $M = 1$ (top), $M = 4$ (middle) and $M = 7$ (bottom), agents learn the strategies that allow them to maintain high acceptance rates and high average payoffs. Notwithstanding, the offered values are higher and fairer if $M$ increases. An average $p$ of 0.2 ($M = 1$) endows Proposers with an average payoff of 0.8, while each Responder keeps 0.2. Oppositely, an average value of $p$ close to 0.7 ($M = 7$) provides the more equalitarian outcome of endowing Proposers with 0.3 and Responders with 0.7. Recall that the sub-game perfect equilibrium always predicts that Proposers would keep almost all the sum and Responders would earn something close 0.

In Fig. 1 we additionally portray the variance of strategies at an individual (middle column) and population level (right column). Initially, propensity values are attributed randomly, sampled from a uniform distribution from 0 to $Q(0)_{max}$. This way, the variance of propensity is initially high, at an agent level. However, the average values of $p$ and $q$ used by each agent are approximately 0.5 for

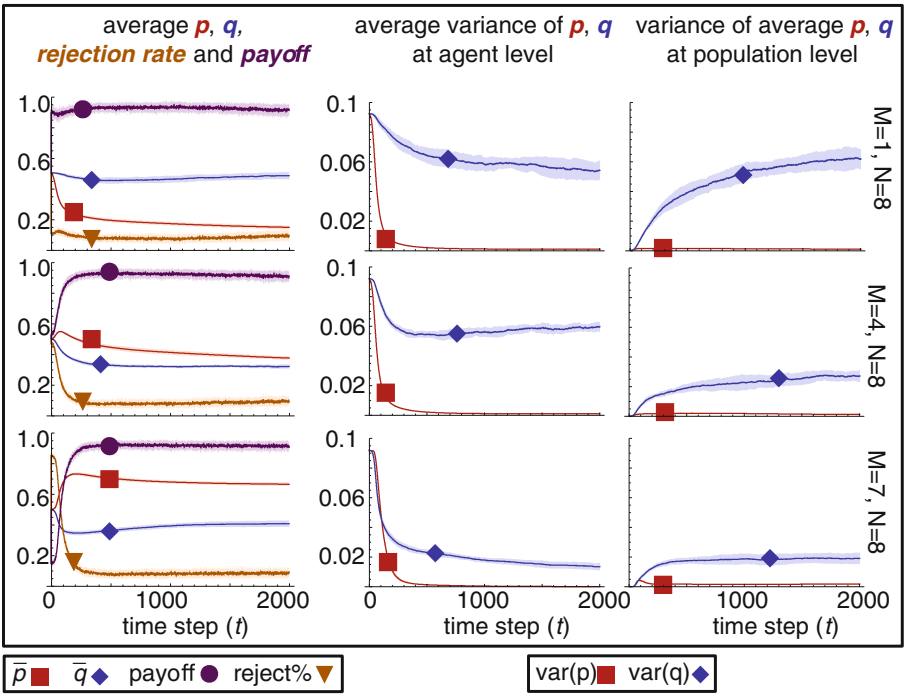


**Fig. 1.** Left column: time series reporting the evolution of average strategies ($\bar{p}$ and $\bar{q}$), average payoff population-wide (*payoff*) and proposals rejection rate (*reject*%). Each plot depicts the average over 100 runs (the corresponding standard deviation, often negligible, is represented by a background shadow), each starting with a random propensity matrix where each entry is sampled from a uniform distribution over the interval $[0, Q(0)_{max}]$. For group size $N = 8$ and for the thresholds $M = 1$, $M = 4$ and $M = 7$, the rejection rate converges to a value near the minimum, thereby, the average payoff in the population approximates the maximum possible. The average strategy values do not inform us about the predictability of agents' actions (the spread of the distribution of individual propensity values) neither about the diversity level of strategies occurring at a population level (spread of average strategies considering all agents), thereby, we present the variance of propensity at an individual (middle column) and population (right column) level. We observe that, as time steps go by, all agents learn to always use (approximately) the same proposal values ($p$), an evidence that stems from the low average variance of propensities both within each agent and across all the population. Contrarily, the variance of the propensity values of $q$ remain high. This variance is considerable lower when $M$ is high, reflecting the larger pressure that is exerted upon $q$. In these specific plots, we assume synchronous propensity updates and offers divided by all the Responders. Other parameters: population size $Z = 100$, granularity $D = 20$, forgetting rate $\lambda = 0.01$, local experimentation rate $\epsilon = 0.01$, total number of time-steps $T = 20000$ $Q(0)_{max} = 20$.

everyone, which results in an initially low variance of average strategies, at a population level. As time goes by (higher values of $t$), all the agents adapt in order to always use the same values of $p$, resulting in a low variance of propensity both at agent and population level. Oppositely, different agents learn to use different ranges of $q$. This is depicted by the high variance of propensity both at an individual and population level: in what concerns $q$, agents are unpredictable and populations are diverse.

We can have a better intuition for the evolving distribution of strategies within a population if we observe a snapshot, for one specific run, of the propensity distribution over the space of possible $p$ and $q$ values. The corresponding results are pictured in Figs. 2 ($M = 1$), 3 ($M = 4$) and 4 ($M = 7$) for time-steps $t = 200$, $t = 500$, $t = 1000$, $t = 3000$, $t = 19000$. Each small square corresponds to a pair $(p, q)$ and a darker square means that more agents have a propensity vector with a high value in that position. Figures 2, 3 and 4 show that, over time, agents learn to use a $p$ value that grows with $M$. Concerning $q$, the learned values have a sizeable variance within the same population. This variance decreases with $M$, an effect already visible in Fig. 1. The reasoning for this result is straightforward: as $M$ increases, a proposal is only accepted if more Responders accept it. In the limiting case of $M = N - 1$, all Responders have to accept an offer in order for it to be accepted by the group, thereby, the pressure for having low acceptance thresholds ($q$) is high. When $M$ is low, a lot of $q$ values in the group of Responders turn to be irrelevant. In this case, the pressure for $q$ values to converge to confined domain is softened.

So far we considered that propensity values are initially attributed at random. This naturally casts doubt on whether populations of initially unfair agents are also able to learn to be fair and adapt their behaviour given different values of $M$. This way, we explicitly consider the effect of initially biased agents. At $t = 0$ we input in each agent a propensity vector that induces them to use a specific strategy with high probability (darker squares in the bottom-left (middle panel) or top-right (bottom panel) corners of each figure at $t = 200$). We consider the two extreme cases of high and low $p$, $q$ values. In the middle panels of Figs. 2, 3 and 4, a lot of propensity is initially placed in the strategies $p = q = 0$, for all agents (extremely unfair agents). In the bottom panels, a lot of propensity is initially placed in the strategies $p = q = 1$ (extremely altruistic agents). We show that, despite this initial bias, agents learn to use approximately the same strategies, in the long run. Moreover, we conclude again that the learned strategies strongly depend on $M$.

Indeed, if we systematically increase $M$, the proposed values rise concomitantly. In Fig. 5 we observe this effect in four different conditions: (*i*) synchronous updates of propensities (Algorithm 1) and payoff divided by all Responders (Eq. 3); (*ii*) synchronous updates of propensities and payoff divided by accepting Responders (Eq. 4); (*iii*) asynchronous updates of propensities (Algorithm 2) and payoff divided by all Responders; (*iv*) asynchronous updates of propensities and payoff divided by accepting Responders. Interestingly, when the payoff is only divided by the accepting Responders, the average values of $q$ and $p$ decrease.

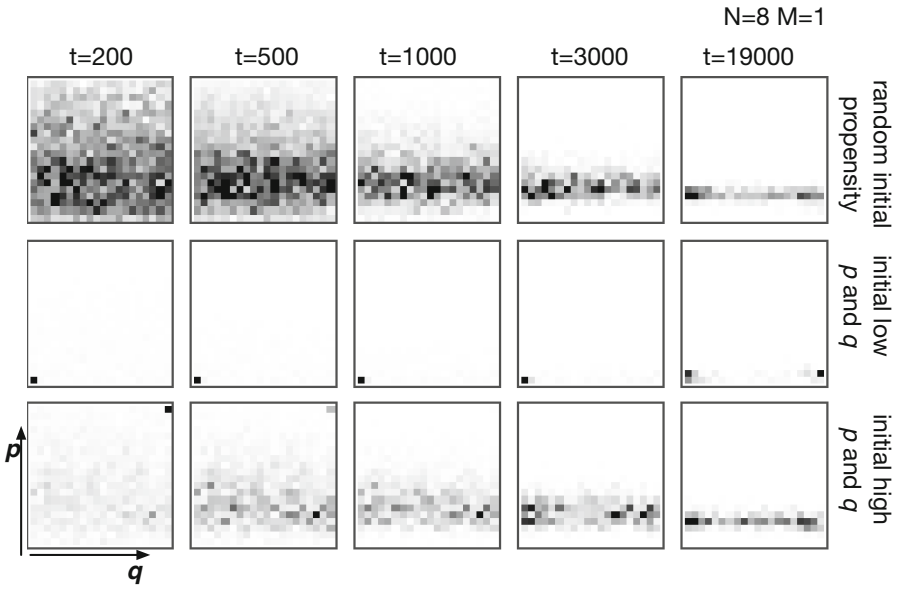


**Fig. 2.** Snapshots of the population composition regarding the average values of $p$ and $q$ to be played given $Q(t)$. Each plot represents the space of all possible combination of $p$ and $q$, assuming that $D = 20$ and thereby, $p$ and $q$ rounded to the closest multiple of $1/20$. We represent the state of the population for five distinct time-steps (from left to right: $t = 200$, $t = 500$, $t = 1000$, $t = 3000$, $t = 19000$) and given three different $Q(0)$ conditions: on top, initial propensity values uniformly distributed; on the middle, initial propensity $Q_{k0}(0) = 50$ and $Q_{k,i\neq 0}(0) = U(0,1)$, where $U = (0,1)$ is a random real sampled uniformly from the interval $[0,1]$; on bottom, initial propensity $Q_{k(D+1)^2-1}(0) = 50$ and $Q_{k,i\neq(D+1)^2-1}(0) = U(0,1)$. Irrespectively of the initial conditions, for $M = 1$ agents learn to use low values of $p$. Each square within the 2D-plots represents a specific combination of $(p, q)$. If the square is darker it means that more individuals use, with high probability, a strategy corresponding to that location. Other parameters: group size $N = 8$, group acceptance threshold $M = 1$, initial propensities maximum $Q(0)_{max} = 50$, population size $Z = 100$, granularity $D = 20$, forgetting rate $\lambda = 0.01$, local experimentation rate $\epsilon = 0.01$, total number of time-steps $T = 20000$.

This result is plausible because, when the number of accepting Responders in a group stands above $M$ and the offer is divided by all the Responders, only the $q$ of those that accepted the proposal has indeed an impact in the obtained payoff; all the agents with a high $q$ receive the same payoff as the accepting agents with low $q$. However, when the payoff is divided by solely the accepting Responders (low $q$), the agents with a high $q$ that individually reject a proposal can be impaired even if the proposal is accepted by the group. This way, the pressure for $q$ to decrease is higher in the condition where proposals are only divided by the accepting Responders. Alongside, the values of $p$ also decrease. Consistently with this hypothesis, when $M$ is higher the difference in both payoff division paradigms is alleviated. On the other hand, there is no significant difference in

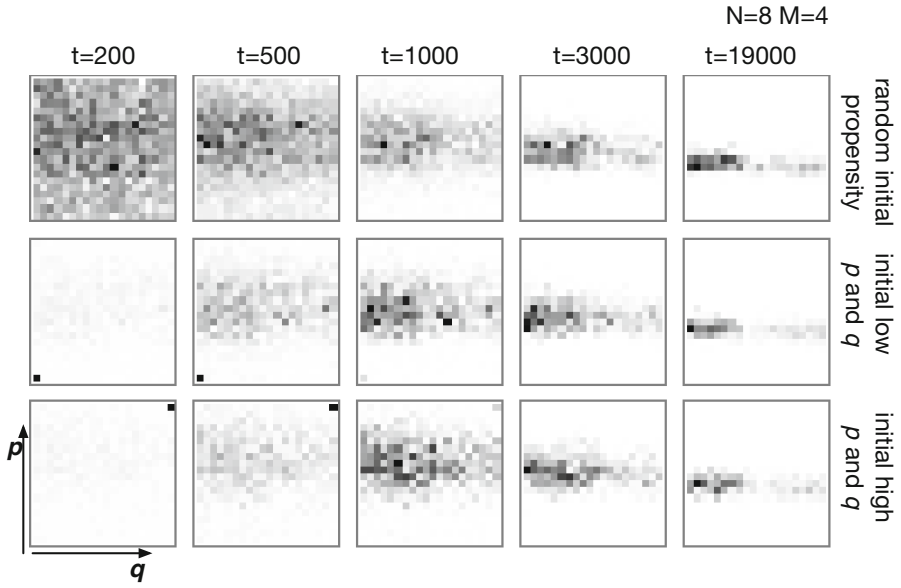


**Fig. 3.** Snapshots of the population composition regarding the average values of $p$ and $q$ to be played given $Q(t)$. For an interpretation of this Fig., please see the caption of Fig. 2. Other parameters: group size $N = 8$ and group acceptance threshold $M = 4$.

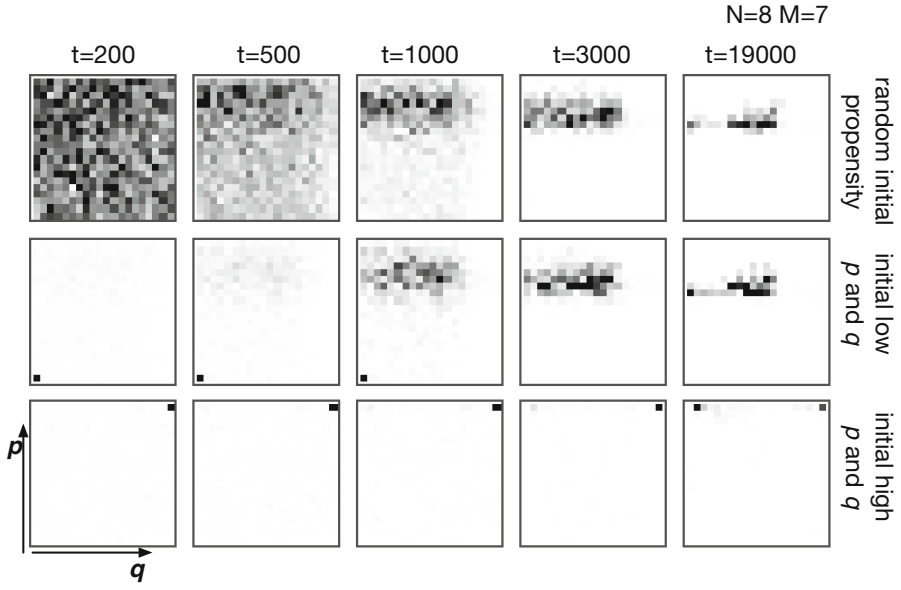


**Fig. 4.** Snapshots of the population composition regarding the average values of $p$ and $q$ to be played given $Q(t)$. For an interpretation of this Fig., please see the caption of Fig. 2. Other parameters: group size $N = 8$ and group acceptance threshold $M = 7$.

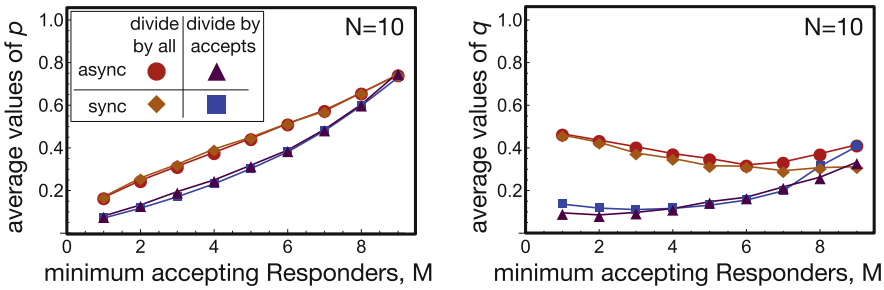


**Fig. 5.** The average values of $p$ and $q$ for group size $N = 10$ with $M$ assuming all possible non-trivial values $1 \leq M \leq N - 1$. Each point in the plot corresponds to a time and ensemble average: $(i)$ time average over the last half of the time-steps, i.e., we wait for a transient time for propensity values to stabilise and $(ii)$ we take the average of 100 runs, each one starting from a random $Q(0)$ propensity matrix. The variance over different runs is negligible. On the left, we represent the average values of proposal and on the right we depict the average threshold acceptance values. In each case, we consider all the combinations of $(i)$ asynchronous or synchronous propensity updates with $(ii)$ payoff divided by all the Responders or payoff only divided by the accepting Responders. Other parameters: population size $Z = 100$, granularity $D = 20$, forgetting rate $\lambda = 0.01$, local experimentation rate $\epsilon = 0.01$, total number of time-steps $T = 10000$, number of runs $R = 100$, initial propensities maximum $Q(0)_{max} = 20$.

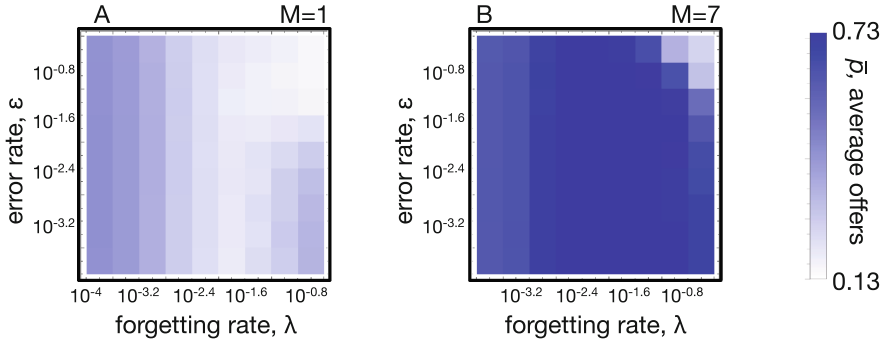


**Fig. 6.** Average values of $p$ and $q$ for different combinations of $\lambda$ (forgetting rate) and $\epsilon$ (local experimentation error). In this case, we assume synchronous propensity updates and offers divided by all the Responders. For all the tested combinations, we always obtain a higher value of $p$ whenever $M$ increases and all other parameters stand fixed. Other parameters: group size $N = 8$, population size $Z = 100$, granularity $D = 20$, total number of time-steps $T = 10000$, number of runs $R = 100$, initial propensities maximum $Q(0)_{max} = 20$.

the learned strategies when considering asynchronous or synchronous propensity updates.

It is noteworthy that the relation between high $M$ and fair proposals remains valid for a wide range of combinations of $\lambda$ (forgetting rate) and $\epsilon$ (local

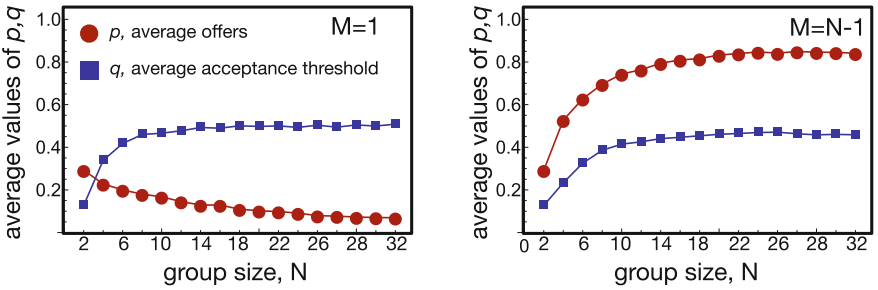


**Fig. 7.** Average values of $p$ and $q$ for different combinations of group sizes, $N$ (2 to 32), and group acceptance rules, $M$ (left panel: $M = 1$ i.e., just a single individual acceptance to render a proposal accepted; right panel: $M = N - 1$, unanimity of individual acceptances to overall accept a proposal). In this case, we assume synchronous propensity updates and offers divided by all the Responders. Other parameters: population size $Z = 100$, granularity $D = 20$, forgetting rate $\lambda = 0.01$, local experimentation error $\epsilon = 0.01$, total number of time-steps $T = 10000$, number of runs $R = 100$, initial propensities maximum $Q(0)_{max} = 20$.

experimentation error) (Fig. 6). We additionally tested for $N = 7$, $M = 1, 3, 6$ and $Z = 20, 30, 50, 200, 300, 500$ and verified that the conclusions regarding the effect of $M$ remain valid for this whole range of population sizes.

Finally, we highlight the effect of group size ($N$) on the average value of proposals made ($\bar{p}$) and proposals willing to be accepted ($\bar{q}$). As Fig. 7 depicts, larger groups induce individuals to rise their average acceptance threshold. As the group of Responders grows and the offers have to be divided between more individuals, the pressure to learn optimal low $q$ values is alleviated. This way, the values of $q$ should increase, on average, approaching the 0.5 barrier that would be predicted if they behaved erratically. Differently, the proposed values exhibit a dependence on the group size that is conditioned on $M$. For mild group acceptance criteria (e.g. $M = 1$), having a big group of Responders is a synonym of having a proposal easily accepted. In these circumstances, Proposers tend to offer less without risking having their proposals rejected, keeping this way more for themselves and exploiting the Responders. Oppositely, when groups agree upon stricter acceptance rules (e.g., $M = 7$), having a big group of Responders means that more people need to be convinced of the advantages of a proposal. This way, Proposers have to adapt, increase the offered values and sacrifice their share in order to have their proposals accepted.

## 5 Discussion and Conclusion

In this work we model an adaptive population with agents interacting through MUG. Agents learn over time following the Roth-Erev reinforcement learning algorithm. This individual learning algorithm was shown to mimic quite well the learning process of humans while playing social dilemmas [8,35]. Indeed, our

main goal is to capture, in a computational model, the role of group acceptance rules and the own group sizes on human behavior. While the role of different group environments is overlooked by an approach that takes all agents as being fully rational, we show that, in the context of learning agents, some particularities of the group setting importantly change the learned behaviors: increasing the minimum number of accepting Responders to obtain group acceptance has the effect of increasing the offered values; the effect of group size depends on the group decision rule in a way that big groups combined with soft group criteria are a fertile ground for selfish Proposers to thrive and, oppositely, big groups that require a large number of Responders to accept a proposal induce Proposers to offer more.

The individual learning model that we implement is close to a trial and error mechanism that individuals may use to successively adapt to the environment, given the feedback provided by their own actions. A different approach implements a system of social learning [38] in which individuals learn by observing the strategies of others and accordingly imitate the strategies perceived as best [31,33,41]. These two learning paradigms (individual and social) can lead to very different outcomes concerning the learned strategies and the long-term behaviour of the agents [4,48]. Interestingly, our results are in line with some of the results obtained in the context of evolutionary game theory and social learning [38]. It is remarkable, however, that an individual learning approach does not rely on information about others' complete strategy set and performances. Agents' learning only requires knowledge about the used strategy and the received payoff. This way, the individual learning method that we employ is suitable to model MUG situations where, reasonably, others' strategies are unknown and the only feedback obtained is the overall group acceptance or rejection.

As stated, we simulate a population of learning agents as a proxy to better understand human behaviour. In AI, algorithms of reinforcement learning are typically implemented in order to equip artificial agents with autonomy and optimality, characteristics often attributed to human intelligence. This way, human behavior is taken as an inspiration to design artificial agents. This work (following others [4,8,22,35,43]) intends to close the loop by experimenting with artificial agents new interactions and behaviors that tentatively allow to gain knowledge about the way that humans act: the emergent behavior of artificial agents is taken as an inspiration to understand human behavior. Interestingly, by telling us something about emergent human behavior, our results can again be used to aid the design of artificial agents that are both efficient and believable when used in human-agent interactions [2,5]. Take the example of automatic negotiation [18,21,34]. What would be the requirements of artificial agents designed to negotiate with a human in an environment that is surely dynamic? Should they behave assuming human rationality and predicting sub-game perfect equilibrium (see Sect. 2.1)? Should they act accordingly with the behaviors that emerge after a learning process? Here we clearly show that the proposal and acceptance threshold of those agents should be implemented as a function of the specific group environment where agents are going to act. These conclusions could hardly be obtained after assuming agents rationality.

Finally, a note on further applications of the game we test with. As Hamilton states, "The theory of many person games may seem to stand to that of two-person games in the relation of sea-sickness to a headache" [15,28]. Indeed, here we see that a multiplayer version of the Ultimatum Game, while still reasonably simple and general, brings attached a set parameters whose effect is certainly not trivial to understand [36,38]. The interaction paradigm that we consider is prevalent in numerous daily situations and human activities, thereby, the study of MUG using different modeling tools and assumptions is both a challenge and opportunity to address stimulating open questions. For instance, how will the group size affect the social pressure on the rejecting Responders? How to manage individual reputations if only the general group verdict is known, rather than individual decisions? What would change if Proposers were allowed to target offers to specific Responders and how would $M$ impact that behavior? How should agents be selected to be Proposers from within a group, given their previous actions? We hope that the adaptive learning agents and multiagent systems community feels tempted to address those (and many other) questions that MUG instigates.

**Acknowledgments.** This research was supported by Fundação para a Ciência e Tecnologia (FCT Portugal) through grants SFRH/BD/94736/2013, PTDC/EEI-SII/5081/2014, PTDC/MAT/STA/3358/2014 and by multi-annual funding of CBMA and INESC-ID (under the projects UID/BIA/04050/2013 and UID/CEC/50021/2013 provided by FCT).

## References

1. Bloembergen, D., Tuyls, K., Hennes, D., Kaisers, M.: Evolutionary dynamics of multi-agent learning: a survey. J. Artif. Intell. Res. **53**, 659–697 (2015)
2. Blount, S.: When social outcomes aren't fair: the effect of causal attributions on preferences. Organ. Behav. Hum. Decis. Process. **63**(2), 131–144 (1995)
3. Bornstein, G., Yaniv, I.: Individual and group behavior in the ultimatum game: are groups more rational players? Exp. Econ. **1**(1), 101–108 (1998)
4. Cimini, G., Sánchez, A.: Learning dynamics explains human behaviour in prisoner's dilemma on networks. J. R. Soc. Interface **11**(94), 20131186 (2014)
5. de Melo, C.M., Carnevale, P., Gratch, J.: The effect of expression of anger and happiness in computer agents on negotiations with humans. In: The 10th International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp. 937–944 (2011)
6. Duch, R., Przepiorka, W., Stevenson, R.: Responsibility attribution for collective decision makers. Am. J. Polit. Sci. **59**(2), 372–389 (2015)
7. Elbittar, A., Gomberg, A., Sour, L.: Group decision-making and voting in ultimatum bargaining: an experimental study. B.E. J. Econ. Anal. Policy **11**(1), 53 (2011)
8. Erev, I., Roth, A.E.: Predicting how people play games: reinforcement learning in experimental games with unique, mixed strategy equilibria. Am. Econ. Rev. **88**, 848–881 (1998)
9. Fischbacher, U., Fong, C.M., Fehr, E.: Fairness, errors and the power of competition. J. Econ. Behav. Organ. **72**(1), 527–545 (2009)

10. Forsythe, R., Horowitz, J.L., Savin, N.E., Sefton, M.: Fairness in simple bargaining experiments. Games Econ. Behav. **6**(3), 347–369 (1994)
11. Fudenberg, D., Levine, D.K.: The Theory of Learning in Games. MIT press, Cambridge (1998)
12. Grosskopf, B.: Reinforcement and directional learning in the ultimatum game with responder competition. Exp. Econ. **6**(2), 141–158 (2003)
13. Güth, W., Schmittberger, R., Schwarze, B.: An experimental analysis of ultimatum bargaining. J. Econ. Behav. Organ. **3**(4), 367–388 (1982)
14. Hagan, J.D., Everts, P.P., Fukui, H., Stempel, J.D.: Foreign policy by coalition: deadlock, compromise, and anarchy. Int. Stud. Rev. **3**(2), 169–216 (2001)
15. Hamilton, W.D.: Innate social aptitudes of man: an approach from evolutionary genetics. In: Fox, R. (ed.) Biosocial Anthropology, pp. 133–155. Wiley, New York (1975)
16. Hoffman, E., McCabe, K., Smith, V.L.: Social distance and other-regarding behavior in dictator games. Am. Econ. Rev. **86**, 653–660 (1996)
17. Iranzo, J., Román, J., Sánchez, A.: The spatial ultimatum game revisited. J. Theor. Biol. **278**(1), 1–10 (2011)
18. Jennings, N.R., Faratin, P., Lomuscio, A.R., Parsons, S., Wooldridge, M.J., Sierra, C.: Automated negotiation: prospects, methods and challenges. Group Decis. Negot. **10**(2), 199–215 (2001)
19. Jing, X., Xie, J.: Group buying: a new mechanism for selling through social interactions. Manage. Sci. **57**(8), 1354–1372 (2011)
20. Kauffman, R.J., Lai, H., Ho, C.-T.: Incentive mechanisms, fairness and participation in online group-buying auctions. Electron. Commer. Res. Appl. **9**(3), 249–262 (2010)
21. Lin, R., Kraus, S.: Can automated agents proficiently negotiate with humans? Commun. ACM **53**(1), 78–88 (2010)
22. Macy, M.W., Flache, A.: Learning dynamics in social dilemmas. Proc. Natl. Acad. Sci. **99**, 7229–7236 (2002)
23. Messick, D.M., Moore, D.A., Bazerman, M.H.: Ultimatum bargaining with a group: underestimating the importance of the decision rule. Organ. Behav. Hum. Decis. Process. **69**(2), 87–101 (1997)
24. Newell, A., Rosenbloom, P.S.: Mechanisms of skill acquisition and the law of practice. Cogn. Skills Acquisition **1**, 1–55 (1981)
25. Nowak, M.A., Page, K.M., Sigmund, K.: Fairness versus reason in the ultimatum game. Science **289**(5485), 1773–1775 (2000)
26. Oosterbeek, H., Sloof, R., Van De Kuilen, G.: Cultural differences in ultimatum game experiments: evidence from a meta-analysis. Exp. Econ. **7**(2), 171–188 (2004)
27. Osborne, M.J.: An Introduction to Game Theory. Oxford University Press, New York (2004)
28. Pacheco, J.M., Santos, F.C., Souza, M.O., Skyrms, B.: Evolutionary dynamics of collective action. In: Chalub, F.A.C.C., Rodrigues, J.F. (eds.) The Mathematics of Darwin's Legacy, pp. 119–138. Springer, Basel (2011)
29. Page, K.M., Nowak, M.A.: Empathy leads to fairness. Bull. Math. Biol. **64**(6), 1101–1116 (2002)
30. Page, K.M., Nowak, M.A., Sigmund, K.: The spatial ultimatum game. Proc. R. Soc. Lond. B Biol. Sci. **267**(1458), 2177–2182 (2000)
31. Pinheiro, F.L., Santos, M.D., Santos, F.C., Pacheco, J.M.: Origin of peer influence in social networks. Phys. Rev. Lett. **112**(9), 098702 (2014)

32. Rand, D.G., Tarnita, C.E., Ohtsuki, H., Nowak, M.A.: Evolution of fairness in the one-shot anonymous ultimatum game. Proc. Natl. Acad. Sci. **110**(7), 2581–2586 (2013)
33. Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M.W., Fogarty, L., Ghirlanda, S., Lillicrap, T., Laland, K.N.: Why copy others? insights from the social learning strategies tournament. Science **328**(5975), 208–213 (2010)
34. Rosenfeld, A., Zuckerman, I., Segal-Halevi, E., Drein, O., Kraus, S.: Negochat-a: a chat-based negotiation agent with bounded rationality. Auton. Agent. Multi-Agent Syst. **30**(1), 60–81 (2016)
35. Roth, A.E., Erev, I.: Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games Econ. Behav. **8**(1), 164–212 (1995)
36. Santos, F.P., Santos, F.C., Melo, F.S., Paiva, A., Pacheco, J.M.: Learning to be fair in multiplayer ultimatum games. In: Proceedings of the 2016 International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1381–1382 (2016)
37. Santos, F.P., Santos, F.C., Paiva, A.: The evolutionary perks of being irrational. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp. 1847–1848 (2015)
38. Santos, F.P., Santos, F.C., Paiva, A., Pacheco, J.M.: Evolutionary dynamics of group fairness. J. Theor. Biol. **378**, 96–102 (2015)
39. Segal-Halevi, E., Hassidim, A., Aumann, Y.: Waste makes haste: bounded time protocols for envy-free cake cutting with free disposal. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp. 901–908 (2015)
40. Sequeira, P., Melo, F.S., Paiva, A.: Emergence of emotional appraisal signals in reinforcement learning agents. Auton. Agents Multi-Agent Syst. **29**(4), 537–568 (2014)
41. Sigmund, K.: The Calculus of Selfishness. Princeton University Press, Princeton (2010)
42. Sinatra, R., Iranzo, J., Gomez-Gardenes, J., Floria, L.M., Latora, V., Moreno, Y.: The ultimatum game in complex networks. J. Stat. Mech. Theory Exp. **2009**(09), P09012 (2009)
43. Skyrms, B.: Signals: Evolution, Learning, and Information. Oxford University Press, Oxford (2010)
44. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. MIT Press, Cambridge (1998)
45. Szolnoki, A., Perc, M., Szabó, G.: Defense mechanisms of empathetic players in the spatial ultimatum game. Phys. Rev. Lett. **109**(7), 078701 (2012)
46. Thaler, R.H.: Anomalies: the ultimatum game. J. Econ. Perspect. **2**, 195–206 (1988)
47. Thorndike, E.L.: Animal intelligence: an experimental study of the associative processes in animals. In: The Psychological Review: Monograph Supplements, (4), i (1898)
48. Van Segbroeck, S., De Jong, S., Nowé, A., Santos, F.C., Lenaerts, T.: Learning to coordinate in complex networks. Adapt. Behav. **18**(5), 416–427 (2010)
49. Van Segbroeck, S., Pacheco, J.M., Lenaerts, T., Santos, F.C.: Emergence of fairness in repeated group interactions. Phys. Rev. Lett. **108**(15), 158104 (2012)
50. Weibull, J.W.: Evolutionary Game Theory. MIT Press, Cambridge (1997)