The Role of Execution Errors in Populations of Ultimatum Bargaining Agents

Fernando P. Santos[1,2], Jorge M. Pacheco[2,3], Ana Paiva[1], and Francisco C. Santos[1,2]

[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. The design of artificial intelligent agents is frequently accomplished by equipping individuals with mechanisms to choose actions that maximise a subjective utility function. This way, the implementation of behavioural errors, that systematically prevent agents from using optimal strategies, often seems baseless. In this paper, we employ an analytical framework to study a population of Proposers and Responders, with conflicting interests, that co-evolve by playing the prototypical Ultimatum Game. This framework allows to consider an arbitrary discretisation of the strategy space, and allows us to describe the dynamical impact of individual mistakes by Responders, on the collective success of this population. Conveniently, this method can be used to analyse other continuous strategy interactions. In the case of Ultimatum Game, we show analytically how seemingly disadvantageous errors empower Responders and become the source of individual and collective long-term success, leading to a fairer distribution of gains. This conclusion remains valid for a wide range of selection pressures, population sizes and mutation rates.

1 Introduction

The attempt to model artificial intelligent agents, revealing human-like behaviour, is often implemented through utility-maximisation heuristics, as *rationality* fits the role of stylised model of human behaviour. When empirical evidence shows that humans systematically deviate from the rational model, explanations suggest the lack of information or computational power. Consequently, the concept of bounded rationality relaxes the strongest assumptions of the pure rational model [15, 27]. Either way, the existence of seemingly irrational decisions is often disadvantageous, if one considers agents in isolation.

Agents are not only intended to act alone in environments, however. Often, they interact in multiagent systems, whose decentralised nature of decision making, huge number of opponents and evolving behaviour stems a complex adaptive system [19]. When agents interact with a static environment, the provided reward functions are well-defined

and the implementation of traditional learning algorithms turns to be feasible. Yet, in the context of the mentioned large-scale multiagent systems, the success of agents strongly depends on the actions employed by the opponents, very much in the same way as we observe in the evolutionary dynamics of social and biological ecosystems [21, 29]. The adaptation of agents, and the learning procedures devised, face important challenges that must be considered [37]. Likewise, the strong considerations about what means to be a rational agent in Artificial Intelligence should be relaxed, or extended. This endeavour can be conveniently achieved through the employment of new tools from, e.g., population dynamics [21] and complex systems research [20], in order to grasp the effects of implementing agents whose strategies, even rational in the context of static environments, may turn to be disadvantageous when successively applied by members of a dynamic population.

In this paper, we present a paradigmatic scenario in which behavioural errors are the source of long-term success. We assume that the goals and strategies of agents are formalised through the famous Ultimatum Game (**UG**) [11], where the conflicting interests of Proposers and Responders likely result in an unfair outcome favouring the former individuals. Additionally, we simulate a finite population of adaptive agents that co-evolve by imitating the best observed actions. We focus on the changes regarding the frequency of agents adopting each strategy, over time. This process of social learning, essentially analogous to the evolution of animal traits in a population, enables us to use the tools of Evolutionary Game Theory (EGT), originally applied in the context of theoretical biology [18]. We start by describing analytically the behavioural outcome in a discretised strategy space of the UG, and in the limit of rare mutations. Additionally, we test the robustness of the results obtained, showing that this analytical approximation remains valid for an arbitrary i) strategy space, ii) selection pressure and iii) mutation rates, via comparison with results from agent-based computer simulations. We shall highlight that this framework, together with a small part of the derived results, were briefly introduced in a short paper [35].

The structure of this paper will continue as follows: Section 2 presents the related work in the scope of the game we experiment with, in the field of EGT and, specifically, in its connections with multiagent learning (MAL). Section 3 will present the methods employed, namely, the analytical model that we employ to study evolution as a stochastic process and the analogous agent-based Monte Carlo simulations. In Section 4, we present the results derived from both methods. Finally, Section 5 is used to provide concluding remarks about the role of execution errors, the nature of its long-term benefits and its relation with the own irrational action.

2 Background

In the present work, we assume that the success of agents is directly derived from the UG [11]. The rules of this game are simple: two players interact in two distinct roles. One is called the *Proposer* and the other is denominated *Responder*. The game is composed by two subgames, one played by each role. First, some amount of a given resource, e.g. money, is conditionally endowed to the Proposer, and this agent must then suggest a division with the Responder. Secondly, the Responder will accept or reject the

offer. The agents divide the money as it was proposed, if the Responder accepts. By rejecting, none of them will get anything. The actions available to the Proposers compose a very large set, constituted by any desired division of the resource. The strategies of the Responders are typically acceptance or rejection, depending on the offer made. We can transform this sequential game in a simultaneous one, if one notes that the strategy of the Responders may be codified *a priori* in a minimum threshold of acceptance [22, 38]. This game condenses a myriad of situations in daily life encounters and in economic interactions. Its use is threefold ideal for the situation we want to analyse; for one way, it allows a simple and objective qualification of utility, as we assume that the success of each agent is uniquely defined by the payoffs earned in the context of this game; for other, this game metaphor is the source of multiple studies that account for an irrational behaviour by human beings [11, 4], considerably hard to justify mathematically; lastly, by being the last round of a bargaining process, the pertinence of this game and its predicted outcome is specially important in artificial intelligence, namely in the design of artificial bargaining agents [16, 14, 17].

The rational behaviour in UG can be defined using a game-theoretical equilibrium analysis, through a simple backward induction. Facing the decision of rejecting (earn 0) or accepting (earn some money), the responder would always prefer to accept any arbitrarily small offer. Secure about this certain acceptance, the Proposer will offer the minimum possible, maximising his own share. Denoting by p the fraction of the resource offered by the Proposer, $p \in [0, 1]$, and by q the acceptance threshold of the Responder, $q \in [0, 1]$, acceptance will occur whenever $p \geq q$ and the *subgame perfect equilibrium* [16] of this game is defined by values of p and q slightly above 0. This outcome is said to be unfair, as it presents a profound inequality between the gains of Proposer and Responder. The strategies of agents that value fairness are characterised by prescribing a more equalitarian outcome: a fair Proposer suggests an offer close to 0.5 and a fair Responder rejects unfair offers, much lower than 0.5 (i.e. $p = 0.5$ and $q = 0.5$).

The rational predictions regarding this game were repeatedly refuted by experimental results. The methods employed to make sense of human decision making and adopted behaviour in UG have, necessarily, to be extended beyond pure rational choice models. The need to disuse optimal methods to model human behaviour, and the relevance of including culturally dependent features, were pointed out in previous works [27, 1]. To overcome the limitation of an optimal rational model, methods related with psychological features and machine learning techniques were proposed [28]. We follow a different path, adopting methods from population ecology, as EGT. EGT was, in the past, successfully used to predict how individual choices may influence the collective dynamics of self-regarding agents, from cells to climate negotiations — see, e.g., [21, 38, 29, 45, 33]. We employ the UG as a game metaphor, without assuming individual or collective rationality. In a social context, EGT describes individuals who revise their strategies through social learning, being influenced by the behaviours and achievements of others [38, 12, 25]. These dynamics of peer-influence allows one to evaluate the extent of the errors and the impact of the own irrational action.

One of the most traditional tools to describe the dynamics of an evolutionary game model is the replicator equation [40]. This equation, justified in a context of trait evo-

lution in biology or cultural evolution across human societies, poses that populations are infinite and evolution will proceed favouring strategies that offer a fitness higher than the average fitness of the population. The fact that replicator equation describes a process of social learning does not prevent it from being convenient in understanding individual learning. A lot of effort has been placed in bridging the gap between replicator dynamics and multiagent learning [2]. Borgers and Sarin showed that there is indeed an equivalence between replicator dynamics and a simple reinforcement learning model (*Cross learning*) [3]. Also, the relationship between Q-learning and replicator dynamics was positively evidenced [43]. It is also important to highlight that EGT is not confined to infinite populations as the replicator equation. The finite nature of real multiagent systems poses the need to consider stochastic effects related with the probabilistic sampling of peers to interact and imitate, a feature that may significantly impact the resulting description of the evolutionary dynamics and the obtained results [13, 33]. This is particularly relevant within multiagent systems research. Thus, in both analytical and numerical computations, we consider evolution as a stochastic process occurring in finite populations [42].

The role of erroneous behaviour during UG encounters was modelled in the past, however in different flavours. Rand et al. studied, both analytically and resorting to experiments, the role of mistakes and stochastic noise in the imitation process of strategies [26]. While that work focus on the role of mutations (or exploration rate) and selection strength (see next section for more details), here of focus on strategic noise, affecting directly the adopted strategies by agents and not the own strategy update process. Notwithstanding, we verify the same general principle: increasing stochasticity, either through high execution errors, low selection strength or small population sizes, has a positive effect on Responders' fitness and overall population fairness. Also in [10] the authors studied errors in executing actions, considering a two-strategy version of UG (the so-called Minigame). The role of errors (in the strategic update process, however) was also studied in the context of multiplayer Ultimatum Games (MUG), both in EGT models [34] and populations with reinforcement learning agents [31]. Next, we present the steps to model the role of execution errors considering large populations of agents, and an arbitrary strategy-space discretisation of UG.

3 Model and Methods

To study the impact of errors in the long-term fitness of agents, we employ two distinct methods. In the first, we describe the prevailing (emergent) behaviours analytically, while resorting to two approximations: we discretise the strategy space of the Ultimatum Game and assume a small mutation (or exploration) probability. These simplifications allow a convenient analytical computation of the most prevailing states, without the need of massive simulations. Notwithstanding, we complement our study with a second method, achieved through simulations that consider the full Ultimatum Game. The simulations are repeated for 100 times (runs) and during 5000 generations. We conclude that the analytical framework proposed provides equivalent results, and may constitute a convenient way of accessing the role of errors in the natural selection of behaviours.

In both methods we consider the existence of two populations (Proposers and Responders) each one composed by Z agents. The adoption of strategies will evolve following an imitation process. The successful individuals will be imitated, thereby, their strategies will prevail in the population. This process of imitation, akin to social learning, fits well with studies that argue for the importance of observing others in the acquisition of behavioural traits [12, 25]. We assume that at each step two agents are chosen, one that will imitate (agent A) and one whose fitness and strategy will serve as model (agent B). The imitation probability will be calculated using a function — $(1 + e^{-\beta(f_B - f_A)})^{-1}$ — that grows monotonously with the fitness difference $f_B - f_A$ [42]. The variable β in the equation above is well-suited to control the selection pressure, allowing to manipulate the extent to which imitation depends on the fitness difference. Whenever $\beta \to 0$, a regime of random drift is attained, in which the imitation occurs irrespectively of the game played, whereas for $\beta \to +\infty$ the imitation will occur deterministically, as even an infinitesimal fitness difference will persuade the imitation of the fitter individual. It is worth to note that, when deciding about imitation, an agent will only observe the fitness of the other agent and the respective strategy; the agents do not have full-information about all interactions of all agents, neither observe the outcome of all individual interactions.

It is also worth to point out that our model copes with fitness, rather than utility. A utility function could vary from agent to agent and could be defined, for instance, to incorporate equality preferences [8], or even a risk-aversion component in the Proposers decision making. Differently, here fitness is uniquely defined by the payoffs of the game, and defines which strategies prevail. The differences between coping with behavioural deviations from rationality through the inclusion of parameters in the utility functions or the study of learning models, is well discussed in [6].

3.1 Analytical framework

Let us assume that a Proposer and a Responder may choose one of S strategies, corresponding to increasing divisions of 1. A Proposer choosing strategy $m \in \{1, 2, ..., S\}$ will offer the corresponding to $p_m = \frac{1}{S}m$ and a Responder choosing strategy $n \in \{1, 2, ..., S\}$ will accept any offer equal or above $q_n = \frac{1}{S}n$. The two-person encounter between a Proposer and a Responder thus yield $1 - p_m$ to the Proposers and p_m to the Responder if the proposal is accepted ($n : q_n \leq p_m$) and 0 to both agents otherwise.

We are concerned with the role of systematic errors in the execution of the desired strategy, namely, by the Responders. The class of these errors should not be mixed with errors implemented in learning procedures, which favour exploration over exploitation, and may naturally provide advantages in deriving optimal policies (as ϵ-greedy methods or *softmax* action selection [39]). Indeed, the errors considered in this paper do not provide a direct feedback to their practitioners and they do not interfere in the social learning procedure. We assume that each Responder with strategy n (and threshold of acceptance q_n) will actually use a threshold of q'_n, calculated as $q'_n = q_n + U(-\epsilon, \epsilon)$, where $U(-\epsilon, \epsilon)$ corresponds to an error sampled from a uniform distribution between $-\epsilon$ and ϵ. Thereby, a Responder (using strategy n) accepts an offer $p_m \in [q_n - \epsilon, q_n + \epsilon]$ with a probability given by $P(q'_n \leq p_m) = P(q_n + U(-\epsilon, \epsilon) \leq p_m) = P(U(-\epsilon, \epsilon) \leq p_m - q_n) = \int_{-\epsilon}^{p_m - q_n} \frac{1}{2\epsilon} d(p_m - q_n) = \frac{p_m - q_n + \epsilon}{2\epsilon}$. The probability of acceptance is 0

if $p_m < q_n - \epsilon$ and is 1 if $p_m \geq q_n + \epsilon$. The resulting payoff of a pair (proposal, acceptance threshold) is, thereby, linearly weighted by the probability of acceptance, considering the execution error (ϵ).

This allows us to compute the average payoffs of each strategy in each interaction, its average fitness and respective transition probabilities (see below). As we assume a well-mixed population, the fitness is given by the average payoff earned when playing with all the agents in the opposite population. Considering the two populations (Proposers and Responders), we say that the population of Proposers is opposite to the population of Responders, and *vice-versa*. Payoff will be defined by encounters between agents from opposite populations and imitation will happen within a population. Thereby, considering the existence of S different strategies in the opposite population of the one from which agent A belongs; denoting k_i as the number of agents using strategy i, in the opposite population of A; and regarding $R_{j,i}$ as the payoff (reward) earned by an agent A using strategy j, against an agent with strategy i (calculated following the rules of UG with execution errors as detailed above), the fitness of agent A is given by

$$f_{A_j} = \sum_{i=1}^{S} \frac{k_i}{Z} R_{j,i} \tag{1}$$

To model the dynamical behaviour of agents when two strategies are present in the population, we adopt the pairwise comparison rule (see [42]), where the imitation probability increases with the fitness difference (see above). Assuming that two agents are randomly sampled from the population in which k_i agents are using strategy i (the remaining are using strategy j), the probability of having ± 1 individual using strategy i is given by

$$T^{\pm}(k_i) = \frac{Z - k_i}{Z} \frac{k_i}{Z - 1} \left(1 + e^{\mp \beta (f_i(\overline{k}_s) - f_j(\overline{k}_s))}\right)^{-1} \tag{2}$$

assuming that in the opposite population the number of agents using another strategy s is $\overline{k}_s$ and that the population size is Z. Note that $\frac{Z-k_i}{Z}$ (and $\frac{k_i}{Z-1}$) represent the sampling probabilities of choosing one agent with strategy $j(i)$ and $\left(1 + e^{\mp \beta (f_i(\overline{k}_s) - f_j(\overline{k}_s))}\right)^{-1}$ translates the imitation probability.

Additionally, with probability μ, a mutation occurs and individuals change their strategy to a random one, exploring a new behaviour regardless the observation of others. The imitation process described above will occur with probability $(1 - \mu)$. As said, if we assume that $\mu \to 0$ [9, 30, 44, 33, 5], we are able to derive analytical conclusions through a simpler apparatus. This simplified limit turns out to be valid over a much wider interval of mutation regimes [44, 33]. Also, while this assumption reduces the random exploration of behaviours, it does not prevent us from considering other stochastic effects, as ϵ, the execution error of Responders. Under this regime in which mutations are extremely rare, a *mutant* strategy will either fixate in the population or will completely vanish [9]. The time between two mutation events is usually so large that the population will always evolve to a monomorphic state (i.e., all agents using the same strategy) before the next mutation occurs. Thus, the dynamics can be approximated by

means of a Markov chain whose states correspond to the different monomorphic states of the populations. This fact allows us to conveniently use Equations (2) in the calculation of transition probabilities, as will be detailed below. Moreover, the time that the populations spend in polymorphic populations is merely transient, thereby disregarded [13, 9].

The transitions between states are described through the fixation probability of every single mutant of strategy i in every resident population of strategy j, that translate how easy is for a strategy originated by a rare mutation, to fixate in a population. A strategy i will fixate in a population composed by $Z-1$ individuals using strategy j with a probability given by [23]

$$\rho_{i\to j}(\overline{k}_s) = \left(\sum_{l=0}^{Z-1}\prod_{k=1}^{l}\frac{T^-(\overline{k}_s)}{T^+(\overline{k}_s)}\right)^{-1} \tag{3}$$

where $\overline{k}_s$ is the number of individuals using strategy s, in the opposite population. Also, while we are calculating the fixation probability in a specific population, the opposite one will remain in the same monomorphic state. This fact allow us to even simplify the calculations [42]. Writing $f_i(\overline{k}_s) - f_j(\overline{k}_s)$ as $\Delta f(\overline{k}_s)$ and noting that $T^-(\overline{k}_s)/T^+(\overline{k}_s) = e^{-\beta\Delta f(\overline{k}_s)}$, Equation (3) reduces to,

$$\rho_{i\to j}(\overline{k}_s) = \frac{1-e^{-\beta\Delta f(\overline{k}_s)}}{1-e^{-Z\beta\Delta f(\overline{k}_s)}} \tag{4}$$

These probabilities define an embedded Markov Chain, governed by the stochastic matrix T, in which $T_{i,j} = \rho_{i\to j}$ defines the fixation probability of a mutant with strategy i in a population with $Z-1$ individuals using strategy j. To calculate π, the stationary distribution of this Markov Process, we compute the normalised eigenvector associated with the eigenvalue 1 of the transposed of T. $\pi_{a,b}$ represents the fraction of time, on average, that is spent when the population of Proposers is using strategy a and the population of Responders is using strategy b. The number of possible states depends on the discretisation chosen, regarding the strategy space considered in Ultimatum Game. If the Proposer and Responder have, each, S available strategies, there are S^2 different monomorphic states. The resulting average fitness is provided by the average fitness of a population in each monomorphic state, weighted by the time spent in that state. Thereby, the average fitness of the population of Proposers is given by $\overline{f} = \sum_{a=1,b=1}^{S}\pi_{a,b}R_{a,b}$ and the average fitness of the population of Responders is given by $\overline{f} = \sum_{a=1,b=1}^{S}\pi_{a,b}R_{b,a}$. Our results refer to $S = 20$. We tested for $S = 10, 20, 30, 40$ and the conclusions remain the same.

3.2 Agent based simulations

The simplifications considered in the previous subsection enable the description of the system as a convenient stochastic process, whose dynamics can be studied without effort. Yet, to know whether the results achieved are sound, we proceeded through agent based simulations, in which agents may choose a continuity of strategies (i.e. $S \to \infty$) and mutations are arbitrary.

We employ a general procedure to simulate evolving agents in the context of EGT. At each time step, a population is picked with probability 0.5. From that population, two agents are chosen (agent A and agent B). The fitness of each agent is calculated by using their strategy against all agents from the opposite population, each with their own strategy. Agent A will then imitate agent B with a probability provided by the sigmoid function — $(1 + e^{-\beta(f_B - f_A)})^{-1}$ — presented in the beginning of this section. With a small probability of μ, imitation will not take place and agent A updates the own strategy to a randomly picked one, between 0 and 1. In biology, this corresponds to a genetic mutation while, in social learning and cultural evolution (and also in typical reinforcement learning algorithms [39]), this mimics the random exploration of behaviours.

The same procedure takes place in the opposite population. When $2Z$ steps of imitation occur, Z in each population, we say that one generation has elapsed. We evolve our system for 5000 generations, and we save the average fitness and average strategy used, for each population. In the beginning, agents start with random strategies, sampled from a uniform distribution between 0 and 1. We repeat the simulation for 50 times, each time starting with random conditions. The results presented (average fitness and average strategy) correspond to a time average over all generations and an ensemble average over all repetitions. In all plays done by the Responders, a noise factor will be added to their base strategy. Thus, the real strategy employed by Responders will correspond to q', their base strategy (q), plus $U(-\epsilon, \epsilon)$, a random value between $-\epsilon$ and ϵ sampled from a uniform distribution in each interaction. Additionally, we followed the same procedure yet assuming a normal distribution with ϵ defining the variance and q defining the mean of the distribution. The same conclusions were obtained, however, the optimal value of ϵ that maximises the Responders' fitness is lower than the one observed with a uniform distribution (but still higher than 0).

4 Results

In this section we report the analytical and numerical results. Anticipating the detailed presentation, we show that the fitness of the Responders will be maximised if they commit a significant execution error, sampled from an interval close to $[-0.3, 0.3]$. Both methodologies are in consonance with this conclusion.

In Figure 1 we show how the average fitness of Proposers and Responders is affected by changing the range of possible execution errors (ϵ) committed by Responders. For different β the conclusion remains equivalent: if the error increases, Responders are endowed with increased fitness. The Proposers are always harmed by the erroneous behaviour of Responders. The *subgame perfect equilibrium* prediction poses that Proposers will earn all the pot, by offering almost nothing to the Responder and assuming an unconditional acceptance by this agent. Yet, if it is assumed that Responders will commit execution errors, which, in the case of heighten the threshold of acceptance may be seen as an irrational behaviour, the Proposers necessarily have to adapt to have their proposals accepted and earn some payoff. This adaptation leads to increased offers (see Figure 3), favouring the average fitness of the Responders. Additionally, we

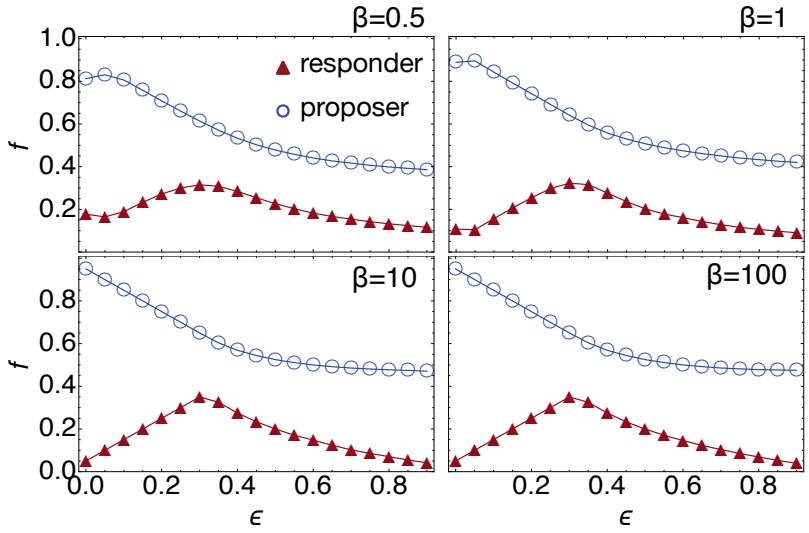

Fig. 1. Analytical results, reporting the average fitness (f) of Proposers (empty circles), average fitness of Responders (filled triangles) for different spans of error committed by the Responders while choosing a strategy, ϵ. It is notorious that it is beneficial for the Responder to behave erroneously, to some extent. If Responders reject irrationally some proposals, the Proposers have to adapt and start to offer more, beneficing, in the long run, the Responders; If Responders reject too much, they will harm themselves and the population of Proposers, as they will waste too much proposals. $Z = 100$, $S = 20$.

note that if the Responders error unreasonably, both Proposers and Responders will be impaired.

One may argue that the results presented in Figure 1 strongly depend on the simplifying assumptions made: the discretisation of strategy space; the assumption of having an equivalence between an average error committed by all agents and the different errors committed individually, within a range; the assumption that most of the time, populations are in a steady monomorphic state, only perturbed by rare mutations. Yet, Figure 2 shows that our conclusions, and the applicability of the analytical model, are more general than one may initially expect. Figure 2 reports the almost exact same results as Figure 1 whereas in this case, they refer to agent based simulations. In these simulations, the assumptions made are disregarded: agents may use any strategy between 0 and 1, each Responder commits a different error within the same range, every time an interaction occurs and no impositions are posed, regarding the time spent in monomorphic states.

The results arguing for an optimal value of error that maximises the fitness achieved by Responders, are also robust for different values of Z (population size), μ (mutation rate) and β (selection pressure). We tested with μ ranging from 0.001 to 0.1 (Figure 2),

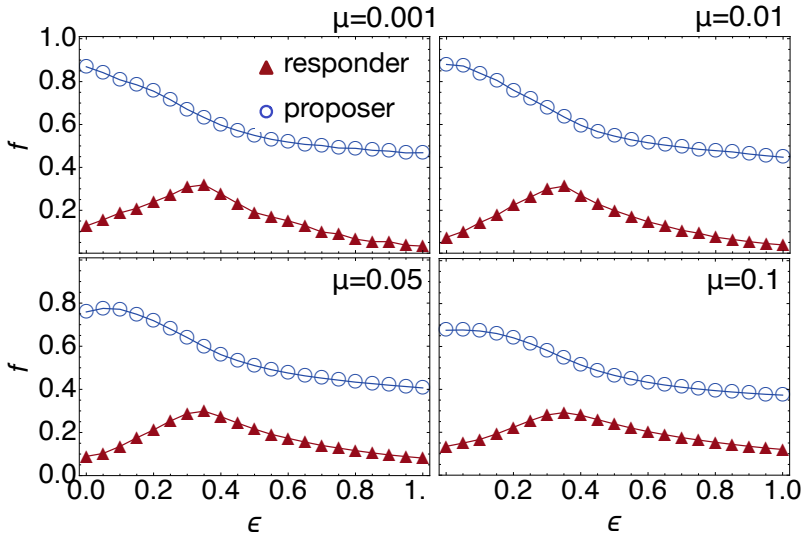

Fig. 2. Results from agent-based computer simulations, reporting the average fitness (f) of Proposers (empty circles), average fitness of Responders (filled triangles) for different spans of error committed by the Responders while choosing a strategy. It is notorious that it is beneficial for the Responder to behave erroneously, to some extent. The results confirm that for a wide range of mutation values (μ), the conclusions regarding the role of execution error (ϵ) in the emergence of fairness remain valid. $Z = 100$, $\beta = 10$

and the conclusions remain valid. Further analytical results regarding β and Z can be accessed in Figures 4 and 5.

In Figure 3 we present the emerging strategies of Proposers (p), regarding the execution error by the Responders (ϵ). The fairer offers made by the Proposers coincide with the highest fitness achieved by the Responders (when ϵ=0.3).

Using both methods (analytical and simulations) it is also possible to assess the impact of β (intensity of selection) and Z (population size) in the emerging average fitness (Figures 4, 5). Again, the analytical and numerical results coincide. Increasing β and Z promotes determinism in the imitation process (see Section 3). Thereby, if the strategy update depends on the fitness difference between agents and no execution errors are considered, the system evolves into a state in which Proposers offer less and Responders accept everything. As an outcome, Proposers keep almost all the payoffs. Even employing a different methodology, these results (regarding the connection between stochasticity and fairness) are in line with the discussion performed in [26, 34].

5 Discussion

In this paper, we apply a novel analytical framework to study UG in a finite population, with an arbitrary state space discretisation. This framework enables the evaluation

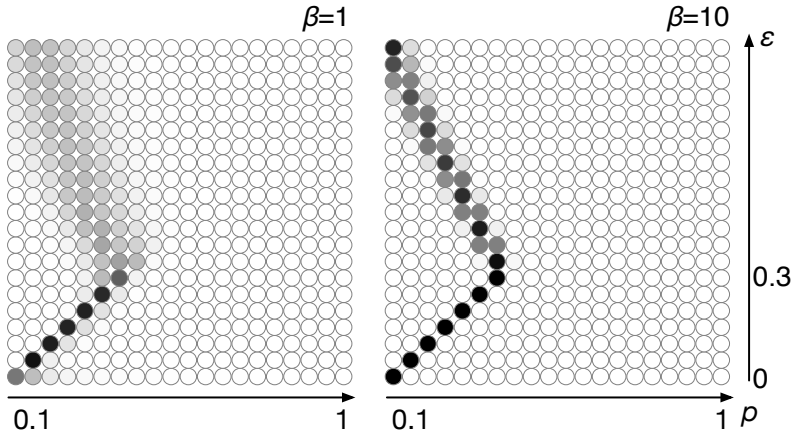

Fig. 3. Prevailing behaviour of Proposers (p) for different execution errors (ϵ) and selection pressure (β), calculated using the proposed analytical framework. Circles coloured using a grayscale represent the stationary distribution over possible *base strategies*, calculated analytically. Darker colours mean that the system spends more time in the corresponding state. It is possible to observe that the Proposers maximise their offers (towards fairer proposals) when $\epsilon = 0.3$. That increased offer coincides with the increase in the average fitness earned by the Responders. $Z = 100$, $S = 20$

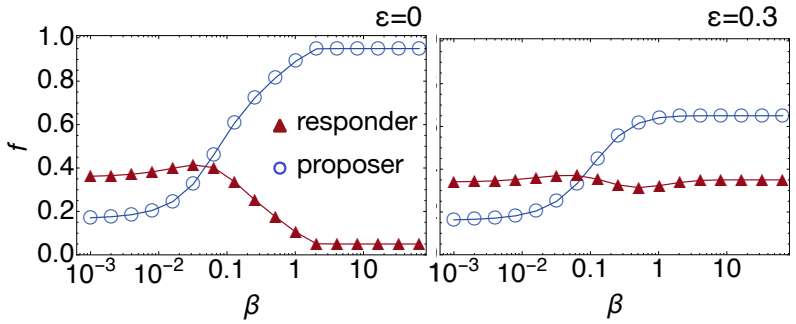

Fig. 4. Analytical results showing the role of selection pressure (β) in the overall fitness (f) of Proposers (empty circles) and Responders (filled triangles), considering two different execution errors by Responders $\epsilon = 0$ and $\epsilon = 0.3$. An increase in β undermines the fitness of Responders. For high intensities of selection, the advantages for Responders of behaving erratically is evident. $Z = 100$, $S = 20$

of stochastic parameters (i.e. selection pressure and different population sizes) in the dynamics of strategy usage in UG. We are able to show, both analytically and through agent-based simulations, how execution errors (ϵ) may promote increased fitness and fairer offers in UG. If the Responders are induced to commit execution errors (which

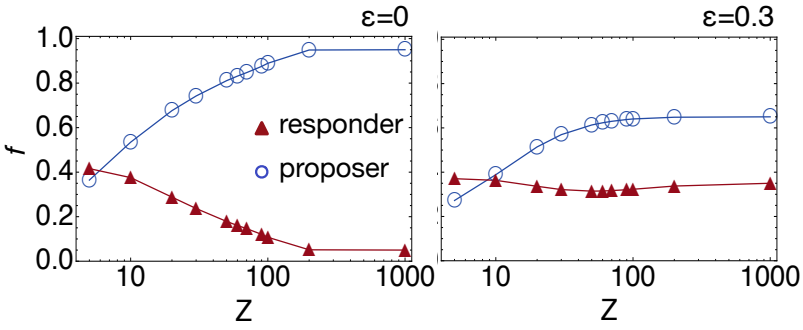

Fig. 5. Analytical results showing the role of different population sizes (Z) in the overall fitness (f) of Proposers (empty circles) and Responders (filled triangles), considering two different execution errors by Responders $\epsilon = 0$ and $\epsilon = 0.3$. An increase in Z undermines the fitness of Responders, similarly to what happened with an increase in β. In fact, high values of Z translate into a population more deterministic, as the fixation of disadvantageous traits by stochastic perturbations turns to be harder. Again, for high population sizes, it is clear the advantage for Responders of behaving erratically. $\beta = 10$, $S = 20$

should not be confused with a mixed strategy, mutation or exploration rate), the seemingly disadvantageous nature of errors turns to be, indeed, an illusion. As Responders error, the Proposers need to adapt and necessarily have to propose generous offers to cover possible errors. Yet, it is also important to understand the extent to which should Responders error. Clearly, being overly erroneous is not beneficial. Other than avoiding the proper adaptation of Responders through the adoption and use of strategies that conduce to fair payoffs, an exaggeration in error span would waste too much proposals, harming both Proposers and Responders. This said, we find an optimal error value in the range $[-0.3, 0.3]$ (i.e., $\epsilon = 0.3$), meaning that, if Responders evolve their base strategy to be close to 0, they would still reject low offers, up to a 0.3 of the total amount help by the Proposer. Despite the plethora of experimental studies in the context of UG, where humans are asked to play this game, a common result is that proposals giving the responder shares below 0.25 are rejected with a very high probability [7], which is interestingly close to the results we obtain.

Finally, we shall highlight that the model we propose avoids the rationality supposition of classical game theory, by assuming that strategies are adopted as the outcome of an evolutionary process of social learning. The relation between our conclusions and eventual results derived from an equilibrium analysis that incorporates noise, as *trembling hand perfect equilibrium* or *quantal response equilibrium*, are naturally interesting. However, by using those game theoretical tools, one would ease the fact that games are often played by agents within adaptive populations, and overall, what seems rational as an individual behaviour may not constitute a good option regarding the collective results [36, 24]. Indeed, in our model, we may identify actions that, even if not rational at an individual scale, turn to be justifiable from an evolutionary point of view.

Execution errors may be seen as a pernicious individual feature that provides collective benefits. Resorting to a multi-level selection mechanism [41], our results may indicate how little pressure evolution may exert to diminish those errors, leading to a plausible argumentation for the natural selection of "erroneous" behaviours, fostered by psychological and emotional factors [4]. Moreover, the fixation of these execution errors is the source of a fairer distribution of payoffs in the UG, as the gains of Responders approximate the gains of the Proposers. As future work, we intend to adapt the proposed framework to analyse the role of errors in group decision making (specifically in multiplayer ultimatum games), where fairness and conflicting interests are paramount [34, 32]

References

1. A. Azaria, A. Richardson, and A. Rosenfeld. Autonomous agents and human cultures in the trust–revenge game. *Auton Agent Multi Agent Syst*, pages 1–20, 2015.
2. D. Bloembergen, K. Tuyls, D. Hennes, and M. Kaisers. Evolutionary dynamics of multi-agent learning: A survey. *J Artif Intell Res*, 53:659–697, 2015.
3. T. Börgers and R. Sarin. Learning through reinforcement and replicator dynamics. *J Econ Theory*, 77(1):1–14, 1997.
4. C. Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, 2003.
5. S. Encarnao, F. P. Santos, F. C. Santos, V. Blass, J. M. Pacheco, and J. Portugali. Paradigm shifts and the interplay between state, business and civil sectors. *R Soc Open Sci*, 2016.
6. I. Erev and A. E. Roth. Maximization, learning, and economic behavior. *Proc Natl Acad Sci USA*, 111:10818–10825, 2014.
7. E. Fehr and U. Fischbacher. The nature of human altruism. *Nature*, 425(6960):785–791, 2003.
8. E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. *Q J Econ*, pages 817–868, 1999.
9. D. Fudenberg and L. A. Imhof. Imitation processes with small mutations. *J Econ Theory*, 131(1):251–262, 2006.
10. J. Gale, K. G. Binmore, and L. Samuelson. Learning to be imperfect: The ultimatum game. *Game Econ Behav*, 8(1):56–90, 1995.
11. W. Güth, R. Schmittberger, and B. Schwarze. An experimental analysis of ultimatum bargaining. *J Econ Behav Organ*, 3(4):367–388, 1982.
12. C. M. Heyes. Social learning in animals: categories and mechanisms. *Biol Rev*, 69(2):207–231, 1994.
13. L. A. Imhof, D. Fudenberg, and M. A. Nowak. Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci USA*, 102(31):10797–10800, 2005.
14. N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, M. J. Wooldridge, and C. Sierra. Automated negotiation: prospects, methods and challenges. *Group Decis Negot*, 10(2):199–215, 2001.
15. D. Kahneman. Maps of bounded rationality: Psychology for behavioral economics. *Am Econ Rev*, pages 1449–1475, 2003.
16. S. Kraus. *Strategic negotiation in multiagent environments*. MIT press, 2001.
17. R. Lin and S. Kraus. Can automated agents proficiently negotiate with humans? *Communications of the ACM*, 53(1):78–88, 2010.
18. J. Maynard-Smith and G. Price. The logic of animal conflict. *Nature*, 246:15, 1973.

19. J. H. Miller and S. E. Page. *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*. Princeton University Press, 2009.
20. M. Mitchell. *Complexity: A guided tour*. Oxford University Press, 2009.
21. M. A. Nowak. *Evolutionary dynamics: exploring the equations of life*. Harvard University Press, 2006.
22. M. A. Nowak, K. M. Page, and K. Sigmund. Fairness versus reason in the ultimatum game. *Science*, 289(5485):1773–1775, 2000.
23. M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. Emergence of cooperation and evolutionary stability in finite populations. *Nature*, 428(6983):646–650, 2004.
24. F. L. Pinheiro, J. M. Pacheco, and F. C. Santos. From local to global dilemmas in social networks. *PLoS One*, 7(2):e32114, 2012.
25. F. L. Pinheiro, M. D. Santos, F. C. Santos, and J. M. Pacheco. Origin of peer influence in social networks. *Phys Rev Lett*, 112(9):098702, 2014.
26. D. G. Rand, C. E. Tarnita, H. Ohtsuki, and M. A. Nowak. Evolution of fairness in the one-shot anonymous ultimatum game. *Proc Natl Acad Sci USA*, 110(7):2581–2586, 2013.
27. A. Rosenfeld and S. Kraus. Modeling agents through bounded rationality theories. *IJCAI'09 Proceedings of the 21st International jont conference on Artifical intelligence*, 9:264–271, 2009.
28. A. Rosenfeld, I. Zuckerman, A. Azaria, and S. Kraus. Combining psychological models with machine learning to better predict people's decisions. *Synthese*, 189(1):81–93, 2012.
29. F. C. Santos and J. M. Pacheco. Risk of collective failure provides an escape from the tragedy of the commons. *Proc Natl Acad Sci USA*, 108(26):10421–10425, 2011.
30. F. C. Santos, J. M. Pacheco, and B. Skyrms. Co-evolution of pre-play signaling and cooperation. *J Theor Biol*, 274(1):30–35, 2011.
31. F. P. Santos, F. C. Santos, F. S. Melo, A. Paiva, and J. M. Pacheco. Dynamics of fairness in groups of autonomous learning agents. In *International Conference on Autonomous Agents and Multiagent Systems*, pages 107–126. Springer, 2016.
32. F. P. Santos, F. C. Santos, F. S. Melo, A. Paiva, and J. M. Pacheco. 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*, pages 1381–1382, 2016.
33. F. P. Santos, F. C. Santos, and J. M. Pacheco. Social norms of cooperation in small-scale societies. *PLoS Comput Biol*, 12(1):e1004709, 2016.
34. F. P. Santos, F. C. Santos, A. Paiva, and J. M. Pacheco. Evolutionary dynamics of group fairness. *J Theor Biol*, 378:96–102, 2015.
35. F. P. Santos, F. C. Santos, A. Paiva, and J. M. Pacheco. Execution errors enable the evolution of fairness in the ultimatum game. In *Proceedings of the 22nd European Conference on Artificial Intelligence (ECAI 2016)*, volume 285, page 1592. IOS Press, 2016.
36. T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.
37. Y. Shoham, R. Powers, and T. Grenager. If multi-agent learning is the answer, what is the question? *Artif Intell*, 171(7):365–377, 2007.
38. K. Sigmund. *The Calculus of Selfishness*. Princeton University Press, 2010.
39. R. S. Sutton and A. G. Barto. *Introduction to reinforcement learning*. MIT Press, 1998.
40. P. D. Taylor and L. B. Jonker. Evolutionary stable strategies and game dynamics. *Math Biosci*, 40(1):145–156, 1978.
41. A. Traulsen and M. A. Nowak. Evolution of cooperation by multilevel selection. *Proc Natl Acad Sci USA*, 103(29):10952–10955, 2006.
42. A. Traulsen, M. A. Nowak, and J. M. Pacheco. Stochastic dynamics of invasion and fixation. *Phys Rev E*, 74(1):011909, 2006.

43. K. Tuyls, K. Verbeeck, and T. Lenaerts. A selection-mutation model for q-learning in multi-agent systems. *AAMAS'03 Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems*, pages 693–700, 2003.
44. S. Van Segbroeck, J. M. Pacheco, T. Lenaerts, and F. C. Santos. Emergence of fairness in repeated group interactions. *Phys Rev Lett*, 108(15):158104, 2012.
45. V. V. Vasconcelos, F. C. Santos, J. M. Pacheco, and S. A. Levin. Climate policies under wealth inequality. *Proc Natl Acad Sci USA*, 111(6):2212–2216, 2014.