Check for updates

Modelling the Dynamics of Subjective Identity in Allocation Games

Extended Abstract

Janvi Chhabra
International Institute of Information Technology, Bangalore
Bangalore, India
janvi.chhabra@iiitb.ac.in

Jayati Deshmukh
International Institute of Information Technology, Bangalore
Bangalore, India
jayati.deshmukh@iiitb.org

Srinath Srinivasa
International Institute of Information Technology, Bangalore
Bangalore, India
sri@iiitb.ac.in

ABSTRACT

Allocation games are zero-sum games that model the distribution of resources among multiple agents. In this paper, we explore the interplay between an *elastic sense of subjective identity* and its impact on notions of fairness in allocation. An elastic sense of identity in agents is known to lead to responsible decision-making in non-cooperative, non-zero-sum games like Prisoners' Dilemma. It thus is a desirable way to model autonomous agents. However, when it comes to allocation, an elastic sense of identity is shown to *exacerbate* inequities in allocation, giving no rational incentive for agents to act fairly towards one another. This leads us to argue that fairness needs to be an innate characteristic of autonomous agency. To illustrate this, we implement the well-known Ultimatum Game between two agents, where their elastic sense of self (represented by γ) and a sense of fairness (represented by τ) are both varied. We study the points at which agents find it no longer rational to identify with the other agent, and uphold their sense of fairness, and vice versa. Such a study also helps us discern the subtle difference between responsibility and fairness in the context of allocation games.

KEYWORDS

Identity; Allocation; Allocation game; Responsibility; Fairness

ACM Reference Format:
Janvi Chhabra, Jayati Deshmukh, and Srinath Srinivasa. 2024. Modelling the Dynamics of Subjective Identity in Allocation Games: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

The allocation of limited resources amongst individuals or groups with competing needs often creates a dilemma especially when conflicting goals, interests, or values are involved. This can be classified under a broad umbrella of responsibility dilemma, where an agent faces a conflict between an individually optimal state and a collectively or socially optimal state. In such dilemmas, classical game theory models agents as rational maximisers i.e., they are modelled to choose actions which maximise their payoff. However, often human societies don't demonstrate such selfish behaviour; humans in such dilemmas do consider factors more than just their personal benefit [10]. Responsible behaviour often emerges in human populations even in times of extreme conflict and oppression [2].

The notion of fairness in agents has been widely studied in resource allocation scenarios [1]. It can be defined as the allocation which an agent perceives as *fair*. Humans facing the dilemma in allocation games demonstrate a preference towards fairness in addition to personal benefit [5]. Fairness can be termed as responsible behaviour since the agents act with the awareness of the impact of their actions on other agents. A recently proposed model called Computational Transcendence (CT) [6] shows that agents get a rational incentive to act responsibly in non-zero sum games, when endowed with an *elastic sense of self*. In this paper, we explore how the concept of CT fares in the context of allocation games. We specifically look at the Ultimatum Game (UG) [7], as the allocation scenario. This game highlights the importance of a sense of fairness and it has also led to research in studying the trade-off between personal and comparative outcomes [8].

2 CT IN ULTIMATUM GAME

Computational Transcendence (CT) [6] models an elastic identity or a sense of self in autonomous agents using which they can identify with external entities like other agents, groups and notions in the system. The sense of self of an agent a is represented as $S(a) = (I_a, d_a, \gamma_a)$ where I_a represents the identity set of the agent consisting of objects it identifies with, d_a is the semantic distance of the agent which denotes the perceived logical distance of an agent to each object in its identity set and γ_a is the transcendence level of the agent which denotes the extent to which it identifies with others. An agent a, with transcendence level γ_a identifies with an object o at distance $d_a(o)$ with an attenuation factor of $\gamma_a^{d_a(o)}$.

Identifying with external entities affects how an agent's internal valuation or *utility* is computed based on external rewards or *payoffs* that are received by different objects in its identity set. For any object $o \in I_a$, let the term $\pi_i(o)$ refer to the payoff obtained by object o in the game or system state i. Given this, the utility derived by agent a in system state i is computed as follows:

$$u_i(a) = \frac{1}{\sum_{\forall o \in I} \gamma_a^{d_a(o)}} \sum_{\forall o \in I} \gamma_a^{d_a(o)} \pi_i(o) \tag{1}$$

In the standard UG between two agents, there exist two roles: Allocator and Recipient. The allocator proposes the allocation of a

This work is licensed under a Creative Commons Attribution International 4.0 License.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

resource between itself and the recipient. Then the recipient can either accept or reject the proposed allocation.

Formally, suppose for a unit resource R, the allocator proposes x for itself and $1 - x$ for the recipient. If the recipient **accepts** the proposed allocation, both allocator and recipient receive the split x and $1 - x$ respectively. On the other hand, if the recipient **rejects** the proposed allocation, then both receive nothing i.e., 0.

The rational strategy using payoff maximization for the allocator is to propose a minimum possible split of resource R to the recipient, and the recipient accepts it because rejecting the allocation results in a payoff of 0.

Transcended agents make the decision based on the expected utility of their choices. Suppose the two transcended agents playing UG are represented as follows: a transcended agent a_1 having a transcendence level of γ and semantic distance d with the other agent a_2. The utility of a_1, when it receives the split of x and a_2 receives the split of $1 - x$, is computed in Equation 2 as follows:

$$util(a_1) = \frac{x + \gamma^d (1 - x)}{1 + \gamma^d} \quad (2)$$

- **If a_1 is the allocator**: It proposes the maximum utility split, i.e., it computes the utility for all the possible splits and proposes the split that gives it the maximum utility.
- **If a_1 is the recipient**: It accepts if the proposed split gives utility $\geq$ to its minimum acceptable utility i.e., it calculates the utility of the proposed split and if it is ≥ 0 it accepts, else it rejects.

We note that a transcended allocator takes almost the whole resource and a transcended recipient accepts the proposed allocation. Since the agents identify with each other, they account for both their and the other agent's payoff. However, in the absence of a notion of fairness, they propose and accept unfair allocations [3]. So we introduce a fairness threshold for transcended agents.

3 INTRODUCING FAIRNESS THRESHOLD

Fairness threshold, (τ) is defined as a threshold below which if an agent receives an allocation, it perceives it to be unfair. Thus, the utility computation accounts for the perceived payoff based on the fairness threshold instead of just the payoff received.

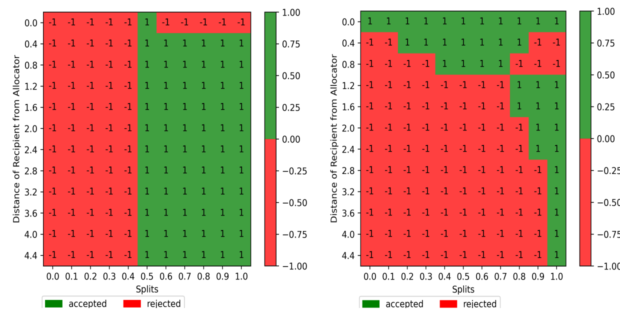
Every agent has a fairness threshold (τ) and it perceives the payoff it receives through the lens of its fairness threshold. An agent a having a fairness threshold of τ_a receives allocation x. Its perceived payoff is computed using a loss-averse sigmoid function. The function $f(x - \tau_a)$, which is a S-shaped function denotes the loss aversion behaviour of the agents [4]. Agents perceive the payoff through the lens of their τ, thus this carries an assumption that other agents' fairness threshold is also the same as their fairness threshold [9]. Hence, utility computation is a combination of perceived payoff and transcended utility computation (Equation 1). For a unit resource, the modified utility computation for agent a incorporating the fairness threshold is described in Equation 3.

$$util(a) = \frac{f(x - \tau) + \gamma^d * f((1 - x) - \tau)}{1 + \gamma^d} \quad (3)$$

We explore different possible ways to represent τ in agents. Following are the two alternatives for representing the fairness threshold in agents:

- **τ as an agent-based characteristic**: Here, τ is a characteristic of the agent itself and it doesn't vary for different objects in the identity set of the agent. This model assumes that the fairness criteria of an agent is universal and same for everyone and thus it does not vary depending on whom the agent is interacting with.
- **τ as an association-based characteristic**: Here, τ of an agent corresponds to each object in its identity set and it depends on how the agent associates with that object. There is a correlation between the semantic distance and fairness threshold for every object. This model assumes that fairness is not a universal characteristic but rather a notion that varies on a case-by-case basis depending on the interaction and perception of individual objects in the identity set.

Figure 1a and 1b represent the acceptable split matrix for the recipient for different splits on the x-axis and varying semantic distance with the allocator on the y-axis. We note that the semantic distance influences the acceptance of split more for association-based τ as compared to agent-based τ. Detailed experiments, results and inferences are discussed in [3].



(a) Agent-based τ (Recipient τ=0.5, γ=0.5)

(b) Association-based τ (Recipient γ=0.5)

4 CONCLUSIONS

The nature of allocation games characteristically differs from other game-theoretic scenarios like Prisoners' Dilemma, since it involves the distribution of a fixed resource, which can be split in multiple ways and yet the collective payoff of all the players always remains the same in all possible game states. Thus the notion of fairness is crucial to be modelled in agents which operate in scenarios representing the allocation games.

In this work, we focused on the Ultimatum Game (UG), which has been widely studied to better understand human behaviour in allocation scenarios. We extended the identity-based model of agents– Computational Transcendence, to allocation games. In this setting, we explored the interplay of the notion of fairness with subjective identity. Our proposed model can be used to simulate agents with diverse behaviours and preferences similar to the variations observed in people across different cultures in the context of allocation games.

ACKNOWLEDGMENTS

We would like to thank the Machine Intelligence and Robotics (MINRO) Center funded by the Government of Karnataka, India and the Center for Internet of Ethical Things (CIET) funded by the Government of Karnataka, India and the World Economic Forum for funding and supporting this work.

REFERENCES

[1] Hamoud S Bin-Obaid and Theodore B Trafalis. 2020. Fairness in resource allocation: Foundation and applications. *Network Algorithms, Data Mining, and Applications: NET, Moscow, Russia, May 2018 8* (2020), 3–18.

[2] Rutger Bregman. 2020. *Humankind: A hopeful history.* Bloomsbury Publishing.

[3] Janvi Chhabra, Jayati Deshmukh, and Srinath Srinivasa. 2024. Modelling the Dynamics of Identity and Fairness in Ultimatum Game. *arXiv preprint arXiv:2401.11881* (2024).

[4] Raymond Dacey. 2003. The S-shaped utility function. *Synthese* 135 (2003), 243–272.

[5] Steven De Jong, Simon Uyttendaele, and Karl Tuyls. 2008. Learning to reach agreement in a continuous ultimatum game. *Journal of Artificial Intelligence Research* 33 (2008), 551–574.

[6] Jayati Deshmukh and Srinath Srinivasa. 2022. Computational Transcendence: Responsibility and agency. *Frontiers in Robotics and AI* 9 (2022).

[7] Werner Güth, Rolf Schmittberger, and Bernd Schwarze. 1982. An experimental analysis of ultimatum bargaining. *Journal of economic behavior & organization* 3, 4 (1982), 367–388.

[8] Richard P Larrick and Sally Blount. 1997. The claiming effect: Why players are more generous in social dilemmas than in ultimatum games. *Journal of Personality and Social Psychology* 72, 4 (1997), 810.

[9] Andrew N Meltzoff. 2005. Imitation and other minds: The "like me" hypothesis. *Perspectives on imitation: From neuroscience to social science* 2 (2005), 55–77.

[10] Amartya K Sen. 1977. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & Public Affairs* (1977), 317–344.