Proceedings of the Second International Workshop on Citizen-Centric Multiagent Systems 2024 (C-MAS 2024)

Co-located with the International Conference on Autonomous Agents and Multiagent Systems (AAMAS'24)

Auckland, New Zealand

Sebastian Stein[1], Archie Chapman[2], Yali Du[3], Behrad Koohy[1], Vahid Yazdanpanah[1], and Pinar Yolum[4]

[1]University of Southampton
[2]University of Queensland
[3]King's College London
[4]Utrecht University

7 May 2024

Welcome to the second edition of C-MAS, where we continue our journey to establish a new focal point for multiagent systems research with a keen emphasis on citizen end-users. Despite the potential of large-scale AI solutions to address societal challenges, users often find themselves on the sidelines, merely providing data and consuming services. C-MAS seeks to challenge this narrative by advocating for alternative approaches that prioritise citizen end-users as active agents with diverse needs and preferences. Through this lens, we aim to establish a new direction for AI systems that is more inclusive, trustworthy, and responsive to the needs of society.

Building upon the foundation laid in C-MAS 2023, this year's event will further explore different aspects of citizen-centric multiagent systems. Throughout the day, attendees will engage with three sessions covering topics such as "Trust and Privacy", "Cooperation and Responsibility", and "Agent-Based Models and Human-Agent Interaction", providing a comprehensive exploration of the new focus point for AI and multiagent systems research.

Further details are available at: https://sites.google.com/view/cmas24

Contents

1 Keynote: What do people really want?

Keynote by Prof. Toby Walsh, University of New South Wales (Sydney)

In multiagent and social choice literature, there's been an extensive analysis of mechanisms for choosing outcomes that satisfy desirable normative properties such as fairness and efficiency. Fairness is interpreted in a number of ways such as envy-freeness or proportionality. But what do people really want? Want can we learn from the psychology and behavioural economics literature?

2 Trust and Privacy

2.1 Distributed Online Life-Long Learning (DOL3) for Multi-agent Trust and Reputation Assessment in E-commerce

Distributed Online Life-Long Learning (DOL3) for Multi-agent Trust and Reputation Assessment in E-commerce

Hariprasauth Ramamoorthy[0009−0004−4922−0319], Shubhankar Gupta[0009−0009−2028−9992], and Suresh Sundaram[0000−0001−6275−0921]

Indian Institute of Science, Bengaluru, Karnataka, India
{hariprasauth,shubhankarg,vssuresh}@iisc.ac.in
https://iisc.ac.in

Abstract. Trust and Reputation Assessment of service providers in citizen-focused environments like e-commerce is vital to maintain the integrity of the interactions among agents. The goals and objectives of both the service provider and service consumer agents are relevant to the goals of the respective citizens (end users). The provider agents often pursue selfish goals that can make the service quality highly volatile, contributing towards the non-stationary nature of the environment. The number of active service providers tends to change over time resulting in an open environment. This necessitates a rapid and continual assessment of the Trust and Reputation. A large number of service providers in the environment require a distributed multi-agent Trust and Reputation assessment. This paper addresses the problem of multi-agent Trust and Reputation Assessment in a non-stationary environment involving transactions between providers and consumers. In this setting, the observer agents carry out the assessment and communicate their assessed trust scores with each other over a network. We propose a novel Distributed Online Life-Long Learning (DOL3) algorithm that involves real-time rapid learning of trust and reputation scores of providers. Each observer carries out an adaptive learning and weighted fusion process combining their own assessment along with that of their neighbour in the communication network. Simulation studies reveal that the state-of-the-art methods, which usually involve training a model to assess an agent's trust and reputation, do not work well in such an environment. The simulation results show that the proposed DOL3 algorithm outperforms these methods and effectively handles the volatility in such environments. From the statistical evaluation, it is evident that DOL3 performs better compared to other models in 90% of the cases.

Keywords: Trust and reputation · Multi-agent systems · E-Commerce.

1 Introduction

Multi-agent systems (MAS) in Distributed Artificial Intelligence (DAI) have the capability to address complex computing problems in Computer Science,

Civil Engineering, Robotics, Economics, etc.; see, for instance [1]. The agents in such an architecture use their knowledge autonomously to decide and act in their environment [2]. One of the major use cases for MAS is in the area of e-commerce, where the agents are distributed in an environment and act autonomously towards their goals, playing various roles like a negotiator, buyer, service provider, consumer, etc [3]. In e-commerce, the agents widely play the role of either a Service Provider or a Service Consumer. The agents act as the representatives of the users at the service provider and service consumer.

In most e-commerce scenarios, the provider would act selfishly to gain the consumer's trust to improve their reputation among the consumers. [4] introduced a novel approach to model this behavior for the service providers by attaching emotional quotients to their interactions. Trust and reputation in such scenarios play a vital role in assisting consumers in identifying the providers to choose from. Adding to the complexity is the noisy data that impacts the way the multi-agents understand the system [5]. Several case studies including those in [6] talk about how malicious sellers deceive and manipulate the viewers. The decentralized marketplace provides better filter and search mechanisms thereby introducing more autonomy for the agents in the interactions [8], further highlighting the importance of Trust and Reputation assessment in such scenarios.

1.1 Contribution

In this paper, we extend the MAS architecture defined in [9] with an observer agent to perform the Trust and Reputation Assessment of service providers. The provider's quality of service can be highly volatile. The incorrect learning during multi-agent interactions leads to a risk that would show infectious growth as agents interact and learn from each other [10].

In this paper, we propose a novel Distributed Online Life-Long Learning (DOL3) framework that involves the online learning of trust and reputation scores of service providers by a set of observers communicating their opinions with each other. Each observer runs the DOL3 algorithm in a decentralized manner. The DOL3 algorithm involves an adaptive online learning framework coupled with a trust fusion process, effectively combining an observer's assessment with its neighboring observers in the interaction network. The online learning process in the DOL3 algorithm is inspired by that of the exponentially weighted online learning forecaster [11]. Simulation results show that DOL3 outperforms the state-of-the-art machine learning assessment methods; such machine learning methods usually involve training a machine learning model to assess an agent's trust and reputation in a stationary environment. On the other hand, owing to its rapid online learning capability, DOL3 deals with the non-stationary environment effectively.

For the simulation studies, different types of social networks have been considered that are essential to understand how the agents are wired to interact with each other and illustrate the social (network) connections among the agents, as stated in [12]. The three networks - Small world, Scale-free, and Regular networks with Homophily are considered during the simulation to understand how

the DOL3 algorithm performs compared to the other methods. To perform the statistical evaluation of these findings, we applied the comparison with real-world data - Movie recommendation system [13], for which the data set was taken from [20].

Simulations involve some recommendation agents becoming malicious in an intermittent fashion. The recommendation system agents are evaluated by the observer agents to help the users get the right list of movies. This environment setup is used to evaluate how the DOL3 algorithm in various network types with different parameters performs compared to that of other state-of-the-art methods and it is evident that the DOL3 algorithm performs better compared to other state-of-the-art models in 90% of the cases.

2 Distributed Online Life-Long Learning (DOL3)

2.1 Problem formulation

The multi-agent architecture considered in this paper involves three types of citizen-centric agents: service providers, consumers, and observers. The edges (connecting lines) between observers and providers indicate that those specific observers have visibility limited to the linked providers. The edges among the observers indicate their neighborhood where the information sharing happens. The consumers can interact with only those providers that they are connected to as per the interaction network. An observer is tasked to do a timely and effective assessment of the providers' quality of service to guide the consumers with the best possible provider.

Let there be N_p service providers, N_o observers, and N_c consumers. Let Ω_i denote the set containing indices of all the providers that are observed by the i^{th} observer as per an interaction network G. Denote Λ_i as the set containing the indices of all the observers that are neighbors to the i^{th} observer as per the interaction network G. Further, let Γ_i be the set of consumers that receive recommendations from the i^{th} observer as per the interaction network G.

It is assumed that the consumers can purchase services one by one w.r.t. interaction count t, with only one consumer per interaction count, i.e., 1^{st} consumer purchases at $t = 1$, 2^{nd} consumer purchases at $t = 2$, and so on, such that the i^{th} consumer purchases only at the interaction counts given by the count sequence: $t_{n,i} = (n-1)\cdot N_c + i$, where $n = 1, 2, \cdots, \lfloor \frac{t}{N_c} \rfloor, \cdots, \infty$, and $i \in [N_c]$. Each service provider j is characterized by a promise quotient $s_j(t) \in [0, 1]$, which is indicative of how good the quality of service provided by the j^{th} provider at the event of its sale at interaction count t is, where $j \in [N_p]$. Further, the service providers have a limited stock of services they sell, characterized by the maximum number of sales/purchases a service provider j can undergo, denoted by n_j^{max}. Let $n_{t,j}$ denote the total number of sales by the j^{th} provider until the interaction count t since it became active. When the sales hit the threshold value n_j^{max} for the j^{th} service provider, the j^{th} provider becomes idle or unavailable to the consumers for the next τ_r interaction steps; this duration serves as the

total number of interaction counts it takes to refill the stock, after which the j^{th} provider becomes active again. The observers are agents that observe the trade between the providers and the consumers, i.e., they observe the promise quotient of a sale/purchase. Based on these observations, the goal of the observers is to recommend high-quality service providers to consumers.

This paper considers a simplified model for the promise quotient capable of simulating various service provider behaviors, ranging from stable to highly volatile that we observe in the e-commerce world [15]. The model is described as follows:

$$s_j(t) = \begin{cases} 1: & \textit{with prob.} \quad p_j(t) \\ 0: & \textit{with prob.} \quad 1 - p_j(t) \end{cases} \tag{1}$$

The DOL3 algorithm consists of three phases: **Periodic Reset Phase:** Assists in frequent forgetting and rapid learning; **Communication Phase:** Shares the scores among the neighbours; **Trust Fusion Phase:** Calculates the weighted trust score based on the scores received; **Learning Phase:** Updates the trust weights using multiplicative exponential weights update scheme.

The details of these phases are explained in Appendix 5.3.

2.2 The DOL3 algorithm

For the i^{th} observer, $\forall i \in [N_o]$, the DOL3 algorithm involves learning the local trust weights $\hat{w}_{ij}(t)$, $\forall j \in \Omega_i$, and social trust weights $\hat{\alpha}^i_{lj}(t)$, $\forall l \in \Lambda_i$ and $\forall j \in \Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)$, which are initialized to 1 at $t = 1$, i.e., $\hat{w}_{ij}(1) = 1$ and $\hat{\alpha}^i_{lj}(1) = 1$. The local trust weight $\hat{w}_{ij}(t)$ represents the degree of trust the i^{th} observer puts on the j^{th} service provider which is directly connected to it as per the interaction network G, $\forall j \in \Omega_i$. On the other hand, the social trust weight $\hat{\alpha}^i_{lj}(t)$ represents the degree of trust the i^{th} observer puts on the l^{th} observer (which is directly connected to the i^{th} observer, $l \in \Lambda_i$) concerning the j^{th} provider's quality of service, where $j \in \Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)$, i.e., j^{th} provider is either directly interacting with the i^{th} observer or the l^{th} observers that are neighbors of the i^{th} observer as per the interaction network G, $\forall l \in \Lambda_i$, or both.

Note that the conditions in equation (2) imply: if j^{th} provider is observed by both the i^{th} and the l^{th} observer, then the social trust weight $\hat{\alpha}^i_{lj}(t+1)$ decreases if there is a mismatch between the i^{th} observer's local trust score $\hat{w}_{ij}(t)$ and the l^{th} observer's local trust score $\hat{w}_{lj}(t)$ since the i^{th} observer would always consider its first-hand observations to be the ground-truth. Whereas, if the j^{th} provider is only observed by the l^{th} observer, the i^{th} observer updates the associated social trust score based on the blind-trust factor $\epsilon^i_{trst,l}$ which is tuned based on how much faith / trust the i^{th} observer can have on its neighboring observers in the interaction network G.

The trust weight $\hat{\alpha}^i_{lj}(t)$ is updated, which indicates how much trust the i^{th} observer has on the l^{th} neighboring observer (as per the interaction network G) for the trust score information on the j^{th} service provider, as follows, $\forall j \in$

$\Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)$, and $\forall l \in \Lambda_i \cup \{i\}$:

$$\hat{\alpha}^i_{lj}(t+1) = \begin{cases} \epsilon^i_{trst,l} : (l = i \wedge j \in \Omega_i) \vee (j \in (\Omega_i \cup \Omega_l) \backslash \Omega_i) \\ (\hat{\alpha}^i_{lj}(t))^\gamma \exp\left(-\eta_\alpha |\hat{w}_{ij}(t) - \hat{w}_{lj}(t)|\right) : \\ \qquad\qquad (l \neq i) \wedge (j \in \Omega_i \cap \Omega_l) \\ 0 : \textit{otherwise} \end{cases} \tag{2}$$

and

$$\alpha^i_{lj}(t+1) = \frac{\hat{\alpha}^i_{lj}(t+1)}{\sum_{l' \in \Lambda_i \cup \{i\}} \hat{\alpha}^i_{l'j}(t+1)} \tag{3}$$

where $\gamma \in (0, 1]$ is the discount factor, and $\eta_\alpha > 0$ is the learning-rate parameter. In equation (2), the first condition represents the case in which either $l = i$ and the j^{th} provider is being observed by i^{th} observer itself, or the j^{th} provider is being observed by the l^{th} observer but not the i^{th} observer. The second condition is valid when $l \neq i$ and the j^{th} provider is being observed by both the i^{th} and j^{th} observers.

3 Performance evaluation

The idea of Agent Reputation and Trust (ART) testbed [22] which is being used for agent trust- and reputation-related technologies is extended further to simulate the real-life citizen-centric scenario of multi-agent systems interaction, which usually includes a lot of complex interactions that result in open and non-stationary environments, which is the main motivation behind developing a simulator to generate uncertainty in the data and include dynamic agents with random behaviors. The evaluation also involves simulating the conditions of how the agents are connected through the social network types - Small world, Random, and Free scale. This ensures that the models are built to scale and work across various types of networks in terms of volume, connectivity, and complexity.

3.1 Simulation evaluation and comparison

The top models from the baseline execution were compared with the DOL3 model in a dynamic environment with multiple Monte Carlo runs. As shown in Fig. 1, the DOL3 and ADST performed better than the rest of the models. The DOL3 algorithm was configured in the simulator by changing the hyper-parameters like n_{reset} to an optimal value along with the discount factor (γ). The heterogeneity of the environment characterized by new providers and the deception of agents characterized by the service quality doesn't impact the speed at which the observers learn the ecosystem. From the various simulation runs, it is evident that DOL3 outperforms the other SOTA models in 90% of the cases.

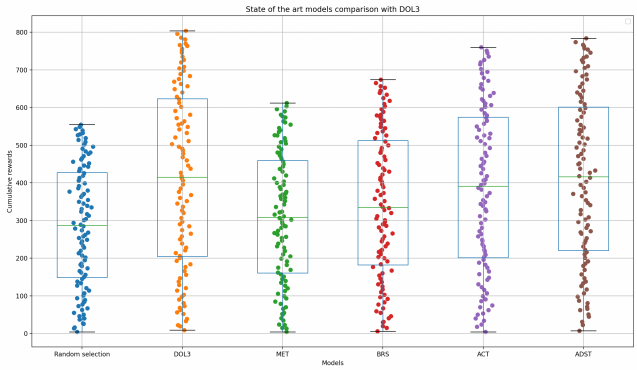


Fig. 1: State-Of-the-art models comparison with DOL3 in Dynamic network

4 Conclusion

With the advancement in e-commerce multi-agent architectures, Trust and Reputation Assessment plays a vital role in ensuring the quality of services. The DOL3 algorithm assists in assessing the trust of the provider and observers in a distributed fashion, where each observer learns to recommend trustworthy service providers to the consumers in real time via DOL3's adaptive online learning architecture. The simulation studies show that DOL3 performs substantially better than machine learning methods like MET, ACT, ADST, SPORAS and HISTOS, owing to its multi-layered online learning coupled with a weighted trust score fusion process and the information sharing among the observers. Further, DOL3's periodic reset phase handles the exploration part of the learning, which takes care of the high volatility in the environment; learned (biased) weights are forgotten and re-initialized after every T_p discrete time-steps. The loss incurred due to such frequent explorations is reduced substantially because of the high convergence rate of DOL3's online learning, owing to the multiplicative exponential weights update scheme. With all the comparisons and statistical evaluations on the real world data, it is evident that DOL3 performs better than the state-of-the-art models in 90% of the cases.

4.1 Limitation and future work

The trade-off between exploration and exploitation in DOL3 needs further investigation. This paper considers all the provider-consumer interactions to be of the same context; DOL3 can be further extended to handle the different or changing contexts scenario. The current problem setting assumes that all the consumers are rational, i.e., they will agree to the observers' recommendations; the problem can be modified further to include irrational consumers as well.

References

1. Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). *Multi-agent systems: A survey*. Ieee Access, 6, 28573-28593.
2. Shamshirband, S., Anuar, N. B., Kiah, M. L. M., & Patel, A. (2013). *An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique*. Engineering Applications of Artificial Intelligence, 26(9), 2105-2127.
3. M. Tomášek and J. Trelová, "An e-commerce applications based on the multi-agent system," 2012 *IEEE 10th International Conference on Emerging eLearning Technologies and Applications (ICETA), Star Lesn*, Slovakia, 2012, pp. 391-394.
4. Jăşcanu, N., Jăşcanu, V., & Nicolau, F. (2007). *A new approach to E-commerce multi-agent systems*. The annals of "Dunarea de Jos "University of Galati. Fascicle III, electrotechnics, electronics, automatic control, informatics, 30, 11-18.
5. Zhang, K., Cao, Q., Sun, F., Wu, Y., Tao, S., Shen, H., & Cheng, X. (2023). *Robust Recommender System: A Survey and Future Directions*. arXiv preprint arXiv:2309.02057.
6. Wu, Q., Sang, Y., Wang, D., & Lu, Z. (2023). *Malicious Selling Strategies in Livestream E-commerce: A Case Study of Alibaba's Taobao and ByteDance's TikTok*. ACM Transactions on Computer-Human Interaction, 30(3), 1-29.
7. S. D. Ramchurn, D. Huynh, and N. R. Jennings, "*Trust in multi-agent systems," The Knowledge Engineering Review*, vol. 19, pp. 1–25, 2004.
8. Tscheke, J., Mr, Attrey, A., Ms, Lesher, M., Ms, Carblanc, A., Ms, & Ferguson, S., Ms (2018). *A Dynamic E-Commerce Landscape: Developments, Trends, and Business Models*. DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION COMMITTEE ON DIGITAL ECONOMY POLICY. https://one.oecd.org/document/DSTI/CDEP(2018)6/en/pdf, pp. 62-64
9. Ehikioya, S. A., & Zhang, C. (2018). *Real-time Multi-Agents Architecture for E-commerce Servers*. Int. J. Networked Distributed Comput., 6(2), 88-98.
10. Chelarescu, P. (2021). *Deception in social learning: A multi-agent reinforcement learning perspective*. arXiv preprint arXiv:2106.05402.
11. Cesa-Bianchi, Nicolo, and Gábor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
12. Dimitri, G. M. (2023). *Is Facebook regionally a small world network?*. arXiv preprint arXiv:2301.04916.
13. Goyani, M., & Chaurasiya, N. (2020). *A review of movie recommendation system: Limitations, Survey and Challenges*. ELCVIA: electronic letters on computer vision and image analysis, 19(3), 0018-37.
14. G. Zacharia & P. Maes, "*Trust management through reputation mechanisms*," Applied Artificial Intelligence, vol. 14, pp. 881-907, 2000.
15. Liu, X. (2007, July). *A multi-agent-based service-oriented architecture for inter-enterprise cooperation system*. In 2007 Second International Conference on Digital Telecommunications (ICDT'07) (pp. 22-22). IEEE.
16. Yu, H. (2014). *Situation-aware trust management in multi-agent systems (Doctoral dissertation)*.
17. Teacy, W. L., Patel, J., Jennings, N. R., & Luck, M. (2006). *Travos: Trust and reputation in the context of inaccurate information sources*. Autonomous Agents and Multi-Agent Systems, 12, 183-198.
18. Jiang, S., Zhang, J., & Ong, Y. S. (2013, May). *An evolutionary model for constructing robust trust networks*. In AAMAS (Vol. 13, pp. 813-820).

19. Wang, N., & Wei, D. (2022). *An Adaptive Dempster-Shafer Theory of evidence Based Trust Model in Multiagent Systems*. Applied Sciences, 12(15), 7633.
20. https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset
21. Huynh, T. D. (2006). *Trust and reputation in open multi-agent systems* (Doctoral dissertation, University of Southampton).
22. Kerr, R., & Cohen, R., 2009. *An Experimental Testbed for Evaluation of Trust and Reputation Systems*. In: Ferrari, E., Li, N., Bertino, E., Karabulut, Y. (eds) Trust Management III. IFIPTM 2009. IFIP Advances in Information and Communication Technology, vol 300. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02056-8_16
23. Masad, D., & Kazil, J. (2015, July). *MESA: an agent-based modeling framework*. In 14th PYTHON in Science Conference (Vol. 2015, pp. 53-60).
24. Kerr, R., & Cohen, R. (2010). *Treet: the trust and reputation experimentation and evaluation testbed*. Electronic Commerce Research, 10, 271-290.
25. Sabater, J., & Sierra, C. (2001, May). *REGRET: reputation in gregarious societies*. In Proceedings of the fifth international conference on Autonomous agents (pp. 194-195).
26. Balakrishnan, V., & Majd, E. (2013). *A comparative analysis of trust models for multi-agent systems*. Lecture Notes on Software Engineering, 1(2), 183.
27. Krampl, J. (2023). of *Thesis: Multi agent model of epidemics with socially. Learning*, 42(8), 1064-1077 (pp. 49-54).

5 Appendix

5.1 Related work

While most researchers rely on contracts for provider-consumer interaction, tracking the transactions and the utility of the corresponding outcome is complex in a large, open, and dynamic environment. In a community of heterogeneous agents where policies define the characteristics of operations, the trust is bounded by the available information. Trust is established based on the past behavior of the agent, and historical events are used computationally to infer or predict future behavior. As stated by [14], there are various basic requirements based on which a trust model can be built, stated as follows:

- Effective trust measure by the trust model
- Capability to handle open MAS
- Robustness against deceptive agents

SPORAS [14], was used in eBay and Amazon by modeling users' trust centrally through rating aggregation. SPORAS does not consider some of the requirements like the domain or context of the environment and past experience in interacting with the provider. ReGret [25] enables each agent to evaluate the reputation by themselves. However, ReGret doesn't take into account the problem of deceptive agents. DOL3 handles the above-stated requirements quite effectively through its multi-layered adaptive online learning of trust scores of the providers and the observers in a decentralized multi-agent architecture with

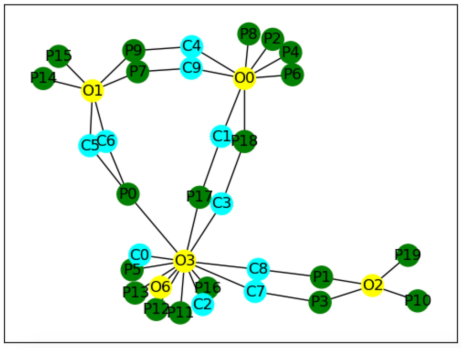


Fig. 2: The communication link among the Observer, Consumer, and Service Provider

information sharing among the observers. The other models like the Trust Computational Model (TCM) and MARSH as specified in [26], use situational and ontological references to compute trust. However, none of them considered the fact of newcomers or changes in the total number of agents in the environment.

As stated in [21], a set of basic requirements (like Interaction trust, Role-based trust, Witness reputation, etc.) must be considered in a Trust and Reputation System. The above set of Trust and Reputation models come up with limitations like SPORAS not considering the social knowledge, HISTOS not having authority on the recommendations, and Beta Reputation System (BRS) having a cold-start problem for new agents entering the environment. Further, some of the evolutionary models that were explored for comparison with DOL3 include: TRAVOS [17], Eigen Trust [21], Actor-Critic-Trust (ACT) [16] - With bootstrap errors, Multiagent Evolutionary Trust model (MET) [18] - Having issues with observers' fairness not considered, Adaptive Dempster-Shafer Theory (ADST) [19] assumes that there is no partial treatment of agents by one another.

5.2 Typical connectivity among the agents

In Fig. 2, the observers (prefixed 'O'), consumers (prefixed 'C'), and providers (prefixed 'P') are labeled and shown as connected over an interaction network to represent their interactions (transactions, observations, and communication). The communication is also restricted to the set of agents as indicated in Fig. 2.

5.3 Phases of DOL3

An iteration of the DOL3 algorithm involves the following phases:

Periodic Reset Phase: The trust weights $\hat{w}_{ij}(t)$, $\forall j \in \Omega_i$, and $\hat{\alpha}^i_{lj}(t)$, $\forall l \in \Lambda_i$ and $\forall j \in \Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)$, are re-initialized to 1 after every T_p interactions. This ensures that the weights do not get biased as the number of interactions increases and can handle the non-stationary nature of the service providers'

behavior. This comes as a consequence of the frequent forgetting along with the rapid learning made possible due to the exponential weights update process in the learning phase.

Communication Phase: as per the interaction network G, i^{th} observer transmits the information $\{t, i, j, \hat{w}_{ij}\}_{\forall j \in \Omega_i}$, and in turn receives the tuples $\{t, l, j, \hat{w}_{lj}\}_{\forall j \in \Omega_l}$ from its neighboring l^{th} observer as per the interaction network G, $\forall l \in \Lambda_i$.

Trust Fusion Phase: the i^{th} observer carries out a weighted fusion of trust weights $\hat{w}_{lj}$ from all its neighboring observers $l \in \Lambda_i$ along with its own trust weight $\hat{w}_{ij}$ for a particular service provider j, $\forall j \in \Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)$, to obtain the i^{th} observer's final trust score of the j^{th} provider, $z_{ij}(t)$, as follows:

$$\hat{z}_{ij}(t) = \sum_{l \in \Lambda_i \cup \{i\}} \alpha^i_{lj}(t) \hat{w}_{lj}(t) \tag{4}$$

$$z_{ij}(t) = \frac{\hat{z}_{ij}(t)}{\sum_{j' \in \Omega_i \cup (\cup_{\forall l \in \Lambda_i} \Omega_l)} \hat{z}_{ij'}(t)} \tag{5}$$

Learning Phase: In this phase, the trust weights are updated using a multiplicative exponential weights update scheme, which is inspired by the exponentially weighted online learning forecaster [11].

The learning phase involves two learning layers; the first one is the local learning layer, in which the i^{th} observer updates the local trust weights for the service providers which are its direct neighbors as per the interaction network G, $\forall j \in \Omega_i$, by utilizing its observations of the purchases, as follows:

$$\hat{w}_{ij}(t+1) = (\hat{w}_{ij}(t))^{\gamma} \exp\left(\eta_w \sum_{k=1}^{k_{t,j}} s_j(t)\right) \tag{6}$$

where $\gamma \in (0, 1]$ is the discount factor, and $\eta_w > 0$ is the learning-rate parameter. Note that $\hat{w}_{ij}(t)$ is indicative of how good the j^{th} providers' quality of service has been as observed by the i^{th} observer.

In the second learning layer, called the social learning layer, the trust weight $\hat{\alpha}^i_{lj}(t)$ is updated.

Further, $\epsilon^i_{trst,l}$ denotes the i^{th} observer's neighbor blind-trust factor for the l^{th} observer, which is equal to 1 for $l = i$ and $\epsilon^i_{trst,l} \in [0, 1]$ for $l \in \Lambda_i$. The blind-trust factor $\epsilon^i_{trst,l}$ represents the degree of blind faith or trust the i^{th} observer put on its neighboring l^{th} observer in the interaction network G. The blind-trust factor $\epsilon^i_{trst,l}$ can be tuned appropriately based on either how much blind trust should be put on a neighboring observer, or to reflect such biases of an observer in real-world scenarios.

5.4 Simulator setup

The simulator architecture is built on the foundation of MESA [23]. The simulator utilizes MESA's basic components, like Agents and Schedulers, to simulate the Agents mentioned in Fig. 5 and their corresponding interactions.

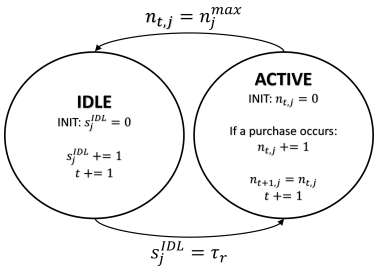


Fig. 3: Active-Idle state switching model for the j^{th} service provider, $\forall j \in [N_p]$; s_j^{IDL} is the step-counter in the idle state, and $n_{t,j}$ is the sales-counter of the j^{th} provider in the active state.

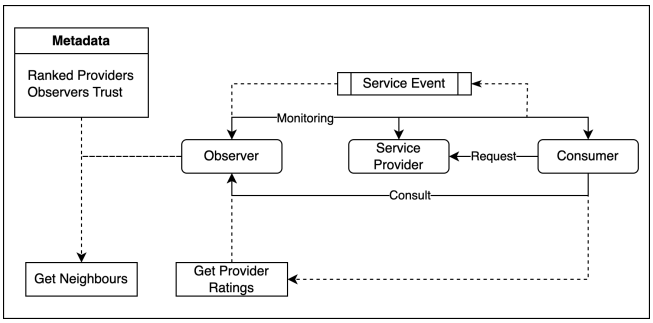


Fig. 4: Simulator architecture

As illustrated in Fig. 4, the main components of the simulator are the components like *Get Neighbours* and *Get Provider Ratings* that help in understanding the network restrictions and weighted fusion rating from all of the observers. There are several configuration capabilities that this architecture provides, allowing the evaluation of the performance of algorithms effectively. Each positive interaction is rewarded with 1, and deceptive interaction is rewarded with 0. The reward is randomized to introduce the non-stationary characteristic of the environment in terms of uncertainty in providers' behavior. The interactions are designed to be sequential per consumer, in the sense that only one consumer interacts with the environment at a time.

One of the important features of the simulator is the interaction restriction among the multiple agents. This paper also shows the behavior of agents when the interactions among the agents are limited to a certain group of agents. As mentioned in 2.1, the consumers can receive services only from a certain set of providers. As described in Section 2.1, each of the providers comes with inventory and restrictions on the number of times it can serve the consumers.

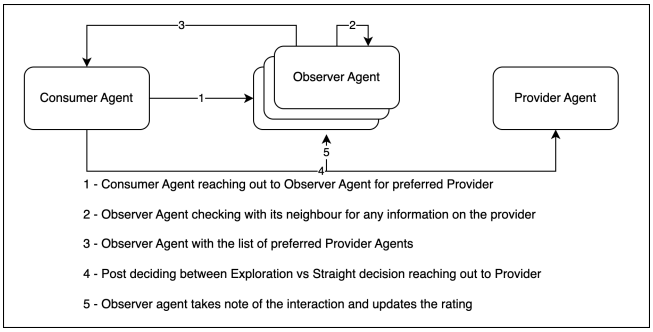


Fig. 5: Simulator sequence

5.5 Model comparisons

Baseline execution The simulation evaluation was done with the baseline version described in 5.6 followed by DOL3 evaluation. Hyperparameters are used that are vital for executing the baseline and then the actual algorithm evaluation. The simulation baseline was set up by configuring the parameters mentioned in Table 1. The randomized baseline starts by assigning random providers irrespective of the scores. The baseline, as well as the algorithm implementation, allows the consumers to either explore or exploit the ranked active providers.

From the Fig. 6a and Fig. 6b, it is clear that the BRS outperforms HISTOS and SPORAS in this simulation environment on both Dynamic and Static networks. From the Decentralised models MDT, ACT, and ADST perform better across the interactions. The simulation is also considered with random observers providing expert opinions on a provider based on past interactions. The baseline clearly illustrates that the openness in the environment with the random behavior of the provider agents impacts the overall reward in the ecosystem, and the learning from the past does not add to improving the reward in the current interactions.

The baseline helps in understanding the level of complexity the open and dynamic environment adds to the ecosystem in building Trust and Reputation. It is also clear that while past learning helps in understanding the agents' behavior, the model built out of the past interaction can not be solely relied on to determine and decide on the agents' behavior for the current interaction. With the network relation in place, the trust needs to be measured from the self-interactions as well as the opinions of the witnesses.

5.6 Baseline for evaluation

Most of the Trust and Reputation Assessment models use SPORAS as the baseline for performance [14], [24]. SPORAS uses the assumption that new users start with little reputation, which builds as the services being provided increase. HISTOS was used for measuring trust in a tightly connected environment [7]. We

extended the baseline to contain some of the state-of-the-art models like ACT, MET, and ADST. The comparisons are done in the order mentioned below on the simulation environment built:

- **Centralised models in Static Network**: Comparison of the models with centralized data update protocol from the references mentioned in the above sections.
- **Centralised models in Dynamic Network**: The centralized models are exposed to the dynamic network where the interaction links change and the number of agents is not consistent.
- **Decentralised models in Dynamic Network**: The decentralized models being exposed to the dynamic network.

The provider agents are ranked based on their reputation or trust scores and recommended to the consumers accordingly. The evaluation is also performed with various parameters. The following categories of baselines were considered:

- **Randomised Baseline**: This baseline randomly assigns reputation or prioritization scores. This lets consumers explore the agents and take a chance to be served by an agent.
- **Expert Opinion Baseline**: This is where the centralized observer methodology comes into the picture. The experts (observers) who know the context and have witnessed the interactions share recommendations about the providers.
- **Start-of-the-Art Models**: The previous State-of-the-Art models were built to measure the Trust and Reputation like that of ADST, ACT, MET, SPORAS, HISTOS, ReGret, MARSH, and TCM.

The evaluation in this simulator consists of a combination of all the above-mentioned baselines. The baselines are customized to fit the problem statement and the characteristics of the environment considered.

Comparison of results We ran a Monte Carlo simulation with 100 steps split between the baseline and DOL3. The experimental result showing the cumulative reward (Sum of all the rewards per iteration) is shown in Fig. 7. It is evident from the results that the baseline is spread on the lower bound of the rewards and is widely spread. However, the DOL3 has very little variance and spread on the upper bound. It is important to notice the variance of DOL3 showcasing that the dynamic environment doesn't impact the quality of the algorithm.

5.7 Statistical validation

The above simulation results help us evaluate the performance of the models against the type of network along with complexity. We applied the same against real-time data of movie recommendation system data set [20]. The recommendation system consisted of consumer agents (users) and service providers (recommenders) along with observers that were connected to represent various social

Table 1: List of hyperparameters used along with description

Hyperparameter		
Variable	Description	Possible Value
N_c	# of Consumers	$\geq$1
N_p	# of Providers	$\geq$1
N_o	# of Observers	≥ 1
N	Total Iterations	$min(100)$
n_{reset}	Every n^{th} reset step	≥ 1
N_{random_stop}	Randomization stops	$min(10)$
$explore$	Explore providers	$True/False$
n^{max}	Maximum provider stock	≥ 1
η	Learning rate - Observer	≥ 1
γ	Discount factor	≥ 0
ϵ	Neighbour Blind-trust	≥ 0
$Observer_{ndepth}$	# of neighbours	$1- \leq (N_c - 1)$

network types like Erdős–Rényi, Watts Strogatz, and Homophily-based networks [27].

Fig. 8a shows how the models perform with the number of interaction counts. The performance or the accuracy is determined by the Root Mean Square Error (RMSE) which is given by the equation:

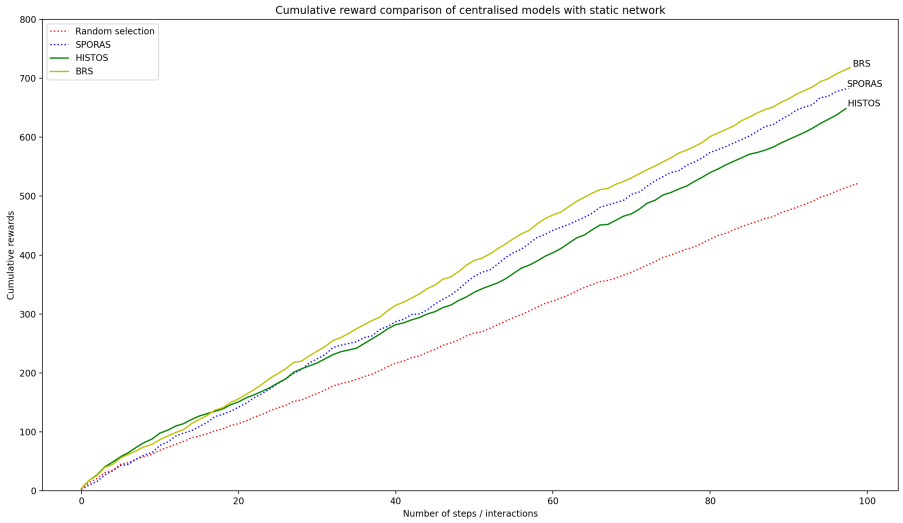
$$RMSE = \sqrt{\frac{1}{N} * (r - \hat{r})^2} \tag{7}$$

where r refers to the actual rating of a movie from the data set and $\hat{r}$ refers to the rating from a recommender.

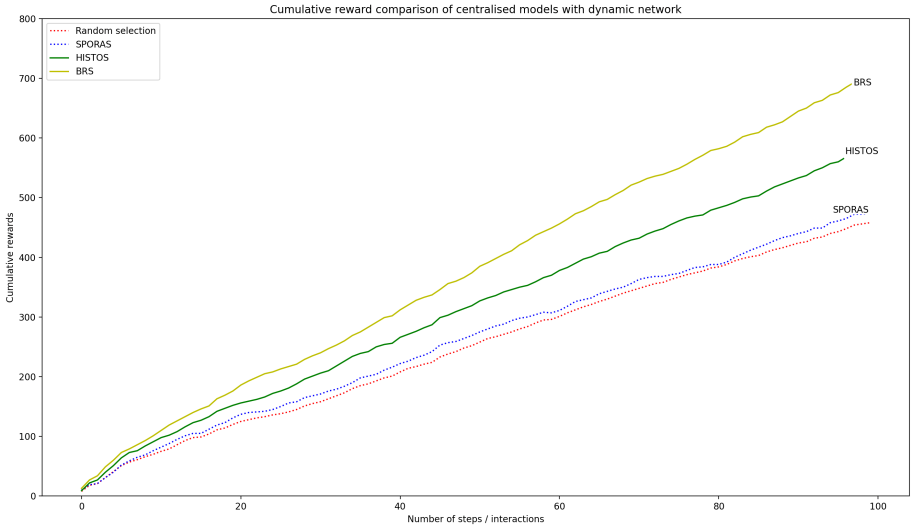
The accuracy is given by

$$accuracy\% = \frac{1}{(1 + RMSE)} * 100 \tag{8}$$

The DOL3 algorithm seems to improve with the larger interaction count compared to that of other models. Similarly, Fig. 8b indicates the performance with the number of malicious agents. We artificially introduced noisy data in the data set to see how the models react when the data is corrupted. We could notice that DOL3 and ADST are more susceptible to noisy data. We could also notice from Fig. 9b that in the network types like that of small world and random, DOL3 performs well.



(a) Static network



(b) Dynamic network

Fig. 6: Centralised models comparison

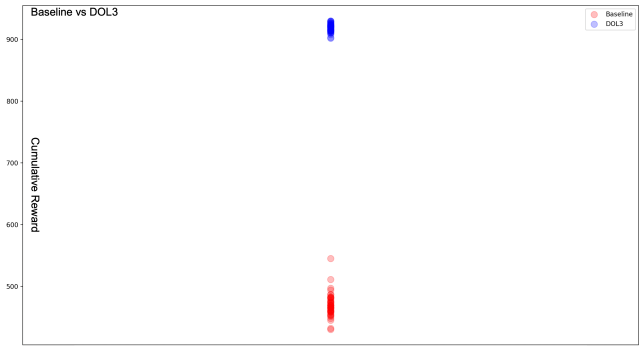
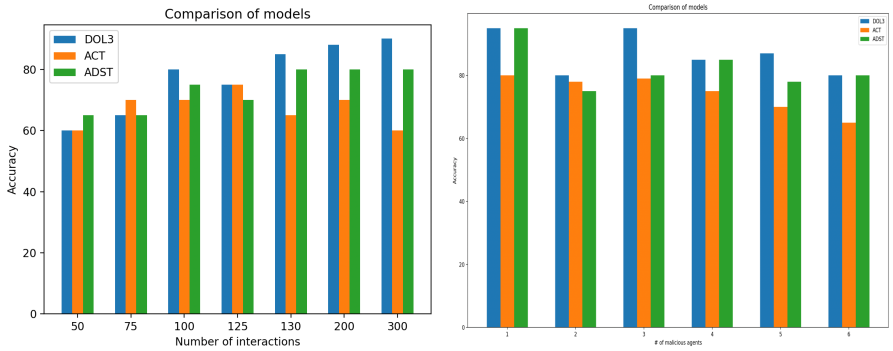
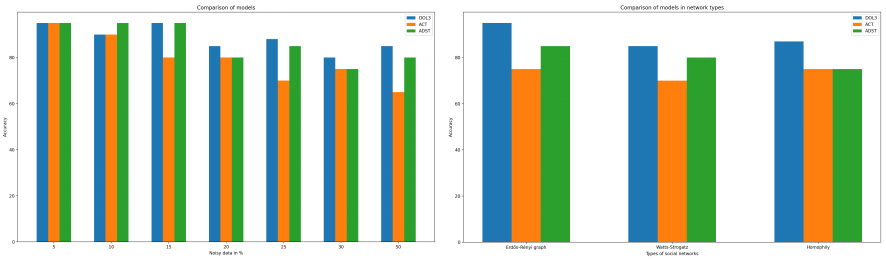


Fig. 7: Simulation results of DOL3 compared with Baseline



(a) Interaction count (b) Number of malicious agents

Fig. 8: Comparison of models with interaction count and malicious agents



(a) Sanity of data (b) Type of network

Fig. 9: Comparison of models based on sanity and type of network

2.2 Resolving Multi-user Privacy Conflicts with Computational Theory of Mind

Resolving Multi-user Privacy Conflicts with Computational Theory of Mind

Emre Erdogan[1][0000−0002−2139−3750], Frank Dignum[1,2][0000−0002−5103−8127], Rineke Verbrugge[3][0000−0003−3829−0106], and Pınar Yolum[1][0000−0001−7848−1834]

[1] Utrecht University, Utrecht, Netherlands
{e.erdogan1,p.yolum}@uu.nl
[2] Umeå University, Umeå, Sweden
dignum@cs.umu.se
[3] University of Groningen, Groningen, Netherlands
l.c.verbrugge@rug.nl

Abstract. Online Social Networks (OSNs) serve as digital platforms for users to share information and build relationships. These networks facilitate the sharing of diverse content, which may disclose personal information about users. Some of these contents pertain to multiple users (such as group pictures), with different privacy expectations. Sharing of such content may lead to *multi-user privacy conflicts*. Decision-making mechanisms are crucial to managing conflicting privacy preferences among users, reducing their effort in conflict resolution. Various mechanisms are proposed in the literature, most of which demand significant computational resources. We propose a novel approach based on computational modeling of *Theory of Mind (ToM)*, the human ability to understand others' mental states (e.g., their beliefs, preferences, goals, etc.), to portray users' privacy expectations. We argue that leveraging computational ToM modeling allows the design of agents capable of accurately capturing users' behavior and reasoning about other agents' privacy understanding, making them effective tools in multi-user privacy conflict management. To illustrate our ideas, we consider a content-sharing scenario and point out potential benefits of using our agent-based computational ToM approach in resolution of privacy conflicts.

Keywords: Multi-user privacy conflict management · Theory of mind · Agent-based model.

1 Introduction

An online social network (OSN) is a digital platform that provides a virtual space for its users to share information, communicate, and build relationships [11]. One of the key characteristics of OSNs is that they allow their users to share various types of content, including photos, videos, and more. The majority of these shared materials have the potential to disclose personal information about the content owner as well as individuals who are associated with it. The users that are all affiliated to a specific piece of content may hold different preferences

about its public visibility since one user might want to make it public whereas another one may prefer to keep it private, leading to potential conflicts, known as *multi-user privacy conflicts* [13].

To automate OSN users' decision-making processes in an accountable manner, we argue that it is important to design agents that can digitally represent users over OSNs and make decisions on their behalf. Ideally, there should be established decision-making mechanisms to manage these situations on the OSNs which can help users to reduce the effort they expend on conflict resolution. Recently, various mechanisms and their associated agent-based solutions have been proposed in the literature. Squicciarini *et al.* [12] propose to use an auction-based mechanism to resolve conflicts, where each user bids for the amount that she sees fit to share or not share a piece of content. Ulusoy and Yolum [15] extend this work to design agents that can learn to bid correctly over time in different types of auctions. Such and Rovatsos [14], Kekulluoglu *et al.* [5], and Filipczuk *et al.* [3] develop systems, where the users' agents negotiate among themselves about various aspects of sharing in order to come to an agreement. In contrast, Kokciyan *et al.* [6] develop a computational argumentation setting, where the agents engage in an argumentation case to decide whether to share or not share a piece of content.

All of these works demonstrate the need and various techniques to resolve multi-user privacy conflicts. However, all of them require substantial computational resources. Put simply, for each piece of content for which a decision needs to be taken among users, one of these mechanisms has to be executed. Conversely, empirical evidence suggests that OSN users manage to deal with real-life privacy conflicts via interpersonal communication [18] and many times by factoring in the privacy understanding of others involved without explicitly involving them into the decision process.

Following this idea, our proposed approach is based on portraying other users' privacy expectations through computational modeling of Theory of Mind (ToM). ToM is generally understood as the human ability to reason about mental content of others such as their beliefs, preferences, intentions, and goals [10,1]. As an important part of social cognition, ToM makes it possible for people to understand and predict others' behaviour. Recent studies around computational ToM models indicate its effectiveness in different settings and tasks [17,2,4]. We argue that with the help of computational ToM modeling, we can design agents that can capture their users' behaviour and reason about other agents accurately, making them an effective tool in multi-user privacy management.

The rest of the paper is organized as follows. First, we begin with a scenario illustrating a privacy conflict unfolding between two users on an OSN. Next, we formally describe preferences and beliefs by using elements from formal logic. Then, we illustrate potential benefits of using our agent-based computational ToM approach in the context of multi-user privacy management. We conclude our work with further research directions.

2 Privacy Management with Theory of Mind

Addressing conflicts in multi-user privacy necessitates a sophisticated understanding of users' privacy preferences. Unlike a simplistic decision where content is merely labeled as "private" or "public", users often adopt a more granular perspective as they can consider multiple audiences when making sharing decisions. For instance, users might choose to share family photos with close friends while withholding them from colleagues, demonstrating respect for their family members' privacy preferences. Context also plays a pivotal role in determining the privacy value of content [9]. Users may opt to limit public access to specific photos, granting permission only under certain conditions or for certain purposes. Moreover, as the number of affected users increases, sharing decisions become more intricate. Various techniques, such as negotiations [14,5] and auctions [12,15], can be employed to resolve conflicts. However, these methods may impose additional burdens on users, including communication overhead and potential disruptions to relationships.

Although we are aware that preserving privacy is much more nuanced than a Boolean representation of "keeping it private" or "making it public", we deliberately simplify our representation of privacy in this paper in order to demonstrate the use of computational ToM. To illustrate our ideas, we consider a content-sharing scenario in which an OSN user, David, wants to publicly share a photo of himself. The photo, denoted as C, also features another user, called Eve, who prefers to keep it private, resulting in a potential conflict.

The OSN users David and Eve are represented by their agents, D and E respectively, who can take certain actions on their behalf. For example, the agents can automatically *share* contents depending on their users' privacy preferences and beliefs about others (e.g., beliefs about others' privacy preferences), *observe* when contents related to their users are being shared by other agents, and *inform* other agents about their own users' privacy preferences. By using these actions, D and E can automatically make sharing decisions and take necessary interactive actions if they observe privacy violations.

2.1 Formal Notation

We denote the privacy preference of an agent X about a piece of content C as $P_X(C, p)$ (i.e., "X prefers C to be p") where p can be either "*private*" or "*public*". We use preferences as the main propositional blocks of our notation which can be associated with negation and conjunction operators as well as belief modalities per agent. The formal notation we use in this paper is mainly based on doxastic logic [8] and loosely based on preference logic [16,7]. To formally represent *preferences and beliefs* of a set of agents $\mathcal{X}$, we use the following language $\mathcal{L}_{PB}^{\mathcal{X}}$ given by the *Backus-Naur* form:

$$\varphi := P_X(C, p) \mid \neg\varphi \mid \varphi \wedge \varphi \mid B_X\varphi$$

Here, C represent contents, p represent privacy preferences (*private* or *public*), and $X \in \mathcal{X}$. For example, $B_D P_E(C, \mathit{private})$ can be read as "the agent

D believes that the agent E prefers the content C to be private". Notice that $B_E B_D P_E(C, private)$, which states that "the agent E believes that the agent D believes that the agent E prefers the content C to be private", is also a member of $\mathcal{L}^{\mathcal{X}}_{PB}$. Formulas with nested epistemic operators allow us to represent agents' higher-order beliefs about other agents beliefs' in a succinct form.

An agent can create beliefs about others and update them over time. The information to realize these could come from different sources, such as observations that lead to certain inferences or explicitly stated information which can be directly adopted. The agent keeps its beliefs along with preferences in its *belief base* and uses them together to make certain decisions for privacy management.

2.2 Examples

Next, we build on the above-mentioned scenario to point out how computational ToM can be beneficial to handle multi-user privacy conflicts.

Considering others' preferences: Our scenario starts with David's agent D, which has (1) $P_D(C, public)$ and can make a sharing decision based on it only. On the other hand, Eve prefers to keep C private, so her agent E has (2) $P_E(C, private)$. If David wants to take Eve's privacy preference about C into account as well as his own before sharing C, D needs to explicitly hold a belief about Eve's preference. This belief may be a correct representation of the actual situation or not. Suppose that D correctly creates the belief (3) $B_D P_E(C, private)$ in its belief base. Table 1 shows both agents' respective belief bases at this stage. Note that each agent can access only its own belief base. Using both David's preference (1) and Eve's assumed preference (3) about C, D can then make a more informed decision, considering also E's expectations, and decide not to make C public on the OSN.

Table 1: D's and E's respective belief bases in the beginning. Based on (1) and (3), D decides not to share C.

D	E
$P_D(C, public)$ (1) $B_D P_E(C, private)$ (3)	$P_E(C, private)$ (2)

Inferring others' beliefs: Suppose D did not accurately model Eve's preference and creates the erroneous belief (4) $B_D P_E(C, public)$. Using (1) and (4), D then decides to share C publicly over the OSN. After observing this action on the OSN, E can infer two pieces of information. First, it infers David's preference about C, resulting in the belief (5) $B_E P_D(C, public)$. Second, it can make another inference about what David believes about Eve's preference. E can interpret the act of sharing as an indication of David's current (incorrect) ToM

Table 2: D's and E's respective belief bases after the sharing action. E creates two new beliefs (5) and (6) accordingly.

D	E
$P_D(C, public)$ (1)	$P_E(C, private)$ (2)
$B_D P_E(C, public)$ (4)	$B_E P_D(C, public)$ (5)
	$B_E B_D P_E(C, public)$ (6)

model of Eve, resulting in the belief (6) $B_E B_D P_E(C, public)$. Table 2 shows both agents' respective belief bases after D's sharing decision.

Dynamically updating (higher-order) beliefs: As the scenario evolves, it is now E's turn to take necessary actions to correct David's ToM model of Eve. Specifically, using the combination of (2), (5), and (6) in Table 2 as a trigger, E can deduce that D needs to be informed about Eve's actual privacy preference about C to make D reconsider its sharing decision (i.e., remove C from the OSN) and inform D accordingly. With this information, D can update its belief base by replacing the erroneous belief (5) $B_D P_E(C, public)$ with the correct belief (4) $B_D P_E(C, private)$. After making sure that D gets this information (for example, by D's acknowledgment), E can update its belief base to correctly represent the actual situation by replacing its higher-order belief (6) $B_E B_D P_E(C, public)$ with (7) $B_E B_D P_E(C, private)$. Table 3 shows both agents' respective belief bases after E's communication with D. Notice that D does not need to inform E again as long as it has the belief (7) in its base.

Table 3: D's and E's respective belief bases after the communication. D replaces (4) with (3) and E replaces (6) with (7).

D	E
$P_D(C, public)$ (1)	$P_E(C, private)$ (2)
~~$B_D P_E(C, public)$ (4)~~	$B_E P_D(C, public)$ (5)
$B_D P_E(C, private)$ (3)	~~$B_E B_D P_E(C, public)$ (6)~~
	$B_E B_D P_E(C, private)$ (7)

Taking proactive actions: So far, we have illustrated the benefits of using computational ToM (with explicitly held beliefs) to resolve an actual privacy conflict. We now change the premise and assume that Eve also prefers C to be public (i.e., (8) $P_E(C, public)$), just like David, as shown in Table 4, but David believes that Eve prefers C to be private (i.e., (3) $B_D P_E(C, private)$). This creates an incorrectly assumed privacy conflict, which stops D to share C on the OSN. Here, E can detect this through ToM. By actively observing D's actions, or in this case the lack of actions, E can hypothesize that D does not

have the correct belief. Building on this observation, E can then proactively inform D about Eve's actual preference. When receiving E's communication, D can use this piece of information to correct its belief about Eve's preference and subsequently reconsider its sharing decision. Table 4 illustrates how both agents' respective belief bases evolve through the example. One can see that each agent holds correct beliefs about the other agent at the end.

Table 4: D's and E's respective belief bases at the end of the alternative scenario. Observing D not making C public helps E to create the beliefs (5) and (7). After E informs D about (8), D replaces (3) with (4) and E replaces (7) with (6).

D	E
$P_D(C, public)$ (1)	$P_E(C, public)$ (8)
~~$B_D P_E(C, private)$ (3)~~	$B_E P_D(C, public)$ (5)
$B_D P_E(C, public)$ (4)	~~$B_E B_D P_E(C, private)$ (7)~~
	$B_E B_D P_E(C, public)$ (6)

3 Conclusion

In this paper, we outline a computational approach based on modeling of theory of mind reasoning for managing privacy conflicts of users of online social networks. By using elements from formal logic, we highlight how computational agents can represent online social networks users' beliefs and privacy preferences explicitly and utilize them to capture the users' content sharing and interacting behaviours. Our concept analysis of a two-user privacy management scenario suggests that computational ToM can be potentially beneficial to agents in various ways in resolving privacy conflicts. As a follow-up work, we aim to develop our ToM-based agent model idea into a more concrete one which is capable of storing, maintaining, and utilizing beliefs by means of proper structures and functions. Also, we will expand the formalization that we use for preferences and beliefs to represent the content-sharing dynamics more realistically (e.g., multiple agents, multiple contents, contexts of contents, etc.). This will enable us to capture real-life, multi-user privacy conflicts observed in online social networks.

Acknowledgments. This research was funded by the Hybrid Intelligence Center, a 10-year programme funded by the Dutch Ministry of Education, Culture and Science through the Netherlands Organisation for Scientific Research, https://hybrid-intelligence-centre.nl, grant number 024.004.022.

References

1. Carruthers, P., Smith, P.K. (eds.): Theories of Theories of Mind. Cambridge University Press (1996)
2. De Weerd, H., Verbrugge, R., Verheij, B.: Higher-order theory of mind is especially useful in unpredictable negotiations. Autonomous Agents and Multi-Agent Systems **36**(2), 30 (2022)
3. Filipczuk, D., Baarslag, T., Gerding, E.H., Schraefel, M.: Automated privacy negotiations with preference uncertainty. Autonomous Agents and Multi-Agent Systems **36**(2), 49 (2022)
4. Gurney, N., Pynadath, D.V.: Robots with theory of mind for humans: A survey. In: 2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN). pp. 993–1000. IEEE (2022)
5. Kekulluoglu, D., Kökciyan, N., Yolum, P.: Preserving privacy as social responsibility in online social networks. ACM Transactions on Internet Technology **18**(4) (2018)
6. Kökciyan, N., Yaglikci, N., Yolum, P.: An argumentation approach for resolving privacy disputes in online social networks. ACM Transactions on Internet Technology **17**(3), 1–22 (2017)
7. Liu, F.: Reasoning about Preference Dynamics, Synthese Library, vol. 354. Springer Science & Business Media (2011)
8. Meyer, J.J.C., Van Der Hoek, W.: Epistemic Logic for AI and Computer Science. No. 41 in Cambridge Tracts in Theoretical Computer Science, Cambridge University Press (2004)
9. Nissenbaum, H.: Privacy as contextual integrity. Washington Law Review **79**, 119 (2004)
10. Premack, D., Woodruff, G.: Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences **1**(4), 515–526 (1978)
11. Schneider, F., Feldmann, A., Krishnamurthy, B., Willinger, W.: Understanding online social network usage from a network perspective. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement. pp. 35–48 (2009)
12. Squicciarini, A.C., Shehab, M., Paci, F.: Collective privacy management in social networks. In: Proceedings of the 18th International World Wide Web Conference. pp. 521–530 (2009)
13. Such, J.M., Criado, N.: Multiparty privacy in social media. Communications of the ACM **61**(8), 74–81 (2018)
14. Such, J.M., Rovatsos, M.: Privacy policy negotiation in social media. ACM Transactions on Autonomous and Adaptive Systems **11**(1), 1–29 (2016)
15. Ulusoy, O., Yolum, P.: PANOLA: A personal assistant for supporting users in preserving privacy. ACM Transactions on Internet Technology **22**(1) (2021)
16. Van Benthem, J., Van Otterloo, S., Roy, O.: Preference logic, conditionals and solution concepts in games. Modality Matters: Twenty-Five Essays in Honour of Krister Segerberg pp. 61–77 (2006)
17. Winfield, A.F.T.: Experiments in artificial theory of mind: From safety to storytelling. Frontiers in Robotics and AI **5**, 75 (2018)
18. Wisniewski, P., Lipford, H., Wilson, D.: Fighting for my space: Coping mechanisms for SNS boundary regulation. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 609–618 (2012)

3 Cooperation and Responsibility

3.1 Selfishness Level Induces Cooperation in Sequential Social Dilemmas

Selfishness Level Induces Cooperation in Sequential Social Dilemmas

Stefan Roesch[1], Stefanos Leonardos[1], and Yali Du[1]

King's College London, London, UK {stefan.roesch, stefanos.leonardos, yali.du}@kcl.ac.uk

Abstract. A key contributor to the success of modern societies is humanity's innate ability to meaningfully cooperate. Game-theoretic reasoning shows, however, that an individual's propensity to cooperation is directly linked with the mechanics of the scenario at hand. Social dilemmas constitute a subset of such scenarios where players are caught in a dichotomy between the decision to cooperate, prioritising collective welfare, or defect, prioritising their own welfare. In this work, we study such games through the lens of 'selfishness level', a standard game-theoretic metric which quantifies the extent to which a game's payoffs incentivize self-directed behaviours. Using this framework, we derive the conditions under which social dilemmas can be resolved and, additionally, produce a first-step towards extending this metric to Markov games. Finally, we present an empirical analysis indicating the positive effects of selfishness level directed mechanisms in such environments.

Keywords: Social Dilemma · Game Theory · Markov Game · Reinforcement Learning · Multi-agent Reinforcement Learning.

1 Introduction

Social dilemmas [9] (SDs) are well studied, and have been the subject of much work in fields such as psychology [3] and sociology [6]. SDs are particularly interesting as they are known to model many real-world coordination problems. A striking example is the case of nuclear weapons proliferation. Here, it is individually rational for a state to maintain a stockpile of nuclear warheads as it serves to deter conflict. However, when multiple states engage in nuclear arms production, the global community becomes endangered by arms races, geopolitical tensions and, accidental use. Ideally, all states should agree to dismantle their nuclear stockpiles, but if any one state were to do so then any opposing, nuclear-armed, states would gain a military advantage. This illustrates that finding solutions to SDs is hard and often requires external mechanisms to align individual incentives with broader societal goals. Sequential social dilemmas (SeqSDs) [8], extend SDs to the Markov game setting and are well known to more accurately represent the complexities of real-world dilemmas. As such they are used as the standard test-bed for mechanisms such as formal contracting [2], social value orientation [10][11], inequity aversion [5][15], and conformity to emergent social norms [14]. In this work we take an interdependence perspective [4], where

individuals are driven not only by 'extrinsic' utilities provided by the environment, but also by an internally realised 'intrinsic' utility, which has recently gained attention in the AI community [10][11][15]. We claim that, in simulated scenarios, extrinsic payoffs can be framed as a miss-specification of objective, requiring some external intervention to align with human values. In this light, we investigate the use of the selfishness level [1] as such an intervention mechanism, studying its effects on SDs and extending the notion to the SeqSD setting, empirically verifying its ability to induce agent cooperation.

2 Selfishness Level & Social Dilemmas

The selfishness level [1] is a scalar metric on the pure Nash equilibria of a normal-form game. Intuitively, a game's selfishness level indicates how much an egotistical player values their own payoff over the collective welfare.

Definition 1 (Selfishness level of a normal-form game [1]). *Given any normal-form game* $G \doteq \{N, \{S_i\}_{i\in N}, \{p_i\}_{i\in N}\}$*, where* N *is a set of players,* S_i*, the strategy space of player* i *and* p_i *the payoff, or utility, function of player* i*, we can induce an altruistic game* $G(\alpha) \doteq \{N, \{S_i\}_{i\in N}, \{r_i\}_{i\in N}\}$ *where,* $r_i(s) \doteq p_i(s) + \alpha SW(s)$*. The selfishness level of a strategic game* G *is:*

$$\alpha_G = \inf_{\alpha}\{\alpha \in R_+ | G \text{ is } \alpha\text{-selfish}\},$$

where, G *is* α*-selfish if, for some* $\alpha \geq 0$*, a pure Nash equilibrium of* $G(\alpha)$ *is a social optimum of* G*.*

SDs [9] are a class of normal-form game which emphasise a dichotomy between individual preferences and the collective good:

	C	D
C	R, R	S, T
D	T, S	P, P

Table 1. Outcome categories in the SD payoff matrix

where R denotes the payoff for mutual cooperation, P mutual defection, S cooperation when an opponent defects and T defection when an opponent cooperates. SDs are further defined by a set of inequalities which prescribe the tensions between individual and collective preferences:

$$R > P,\ R > S,\ 2R > T + S, \tag{1}$$

$$T > R \text{ (greed) or } P > S \text{ (fear)}. \tag{2}$$

The inequalities in 1 work to establish mutual cooperation as the unique, stable, social optimum and the inequalities in 2 dictate the modality of the SD (e.g., when both inequalities in 2 are satisfied, the resultant game is a prisoner's dilemma).

3 Resolving Social Dilemmas

Examining SDs through the lens of selfishness level highlights some interesting properties. We delegate proofs for all theorems to Appendix A.

Theorem 1. *The selfishness level of a SD is*

$$\alpha_G = \begin{cases} 0 & \text{if } T \leq R, \\ \frac{T-R}{2R-T} & \text{if } T > R. \end{cases} \tag{3}$$

Equation 3 shows that when players are troubled only by an equilibrium selection problem (i.e., when G is a stag hunt dilemma), $\alpha_G = 0$. Conversely, when $\alpha_G > 0$ (i.e. a prisoner's or chicken dilemma), the game is not naturally conducive to cooperation. Intuitively, the selfishness level is directly linked with T and R - the greater the value of α_G, the higher the incentive to deviate from mutual cooperation, and vice-versa. As such, the selfishness level formally quantifies the magnitude of the intervention required to realise cooperation.

Here, we investigate how the selfishness level can be used in the design of intrinsic payoff mechanisms to align the players' preferences towards mutual cooperation. Our analysis shows that, in chicken and prisoner's dilemmas, the resultant selfishness level modified payoffs can be relieved of any individual-group tensions, that is, neither inequality in 2 holds.

Theorem 2. *Given a SD G, let $T > R$ and $P \leq 0$ (a chicken dilemma). $G(\alpha)$ is always resolved when $\alpha = \alpha_G$.*

The result of Theorem 2 reflects the fact that the selfishness level works only to alleviate the burden of greed. As chicken dilemmas are troubled only by greed it is natural that the altruistic game induced by α_G, $G(\alpha_G)$ is free of any dilemma.

Theorem 3. *Given a SD G, let $T > R$ and $P > 0$ (a prisoner's dilemma). $G(\alpha)$ is resolved when $\alpha = \alpha_G$ and $P \leq T - R$*

Theorem 3 mirrors Theorem 2. If the personal benefit of exploitation is the driving force behind a player's willingness to defect then a selfishness level modification of payoffs is able to completely resolve the dilemma. Conversely, if $P > T - R$, $G(\alpha_G)$ is a stag hunt.

3.1 Extending to Markov Games

We present here a 'first step' towards the highly non-trivial goal of theorising the selfishness level in the Markov game setting starting with two-player *sequential social dilemmas* (SeqSDs).

Definition 2 (Sequential Social Dilemma [8]). *SeqSDs are characterised by the presence of* critical states $S_c \subseteq S$. *Each $s_c \in S_c$ induces a sub-game such that players' preferences can be expressed as a social dilemma. This can be more easily intuited through table 2.*

	π^C	π^D
π^C	$R(s_c), R(s_c)$	$S(s_c), T(s_c)$
π^D	$T(s_c), S(s_c)$	$P(s_c), P(s_c)$

Table 2. Empirical payoff matrix for $s_c \in S_c \subseteq S$.

where, $R(s_c) \doteq V_i^{\pi_i^C, \pi_{-i}^C}(s_c)$*, and* $T(s_c), S(s_c)$ *and,* $P(s_c)$ *are defined analogously.*

Our extension is defined via the *altruistic Markov game*, which is analogous to the normal-form game presented in definition 1.

Definition 3 (Altruistic Markov Game). *Given a Markov game,* $\mathcal{M}$*, we can induce an altruistic Markov game (cf. Definition 1)*

$$\mathcal{M}(\alpha) := \{N, S, \{A^i\}_{i\in\{1,\dots,N\}}, P, \{\lambda^i\}_{\{i\in\{1,\dots,N\}}, \gamma\}$$

where $\lambda_i(s, a, s') := R_i(s, a, s') + \alpha(\sum_{j\in N} R_j(s, a, s'))$.

We now define a scalar-valued selfishness level for the SeqSD.

Definition 4 (Selfishness Level of Two-Player SeqSDs). *Consider the special case of altruistic Markov games where the host game is a two-player SeqSD. Each* $s_c \in S_c$ *can be considered as a normal-form sub-game. Given this, we construct the set*

$$\vec{\alpha} \doteq \{\alpha_{s_c} | \alpha_{s_c} = \frac{T(s_c) - R(s_c)}{2R(s_c) - T(s_c)} \forall s_c \in S_c\},$$

and define the selfishness level of the SeqSD as

$$\Gamma = \max_{s_c} \vec{\alpha}.$$

It is known that, if for some $\alpha \geq 0$ a social optimum of $G(\alpha)$ is Nash, then it remains as such for every $\beta \geq \alpha$ [1]. I.e., for an s_c with selfishness level α_{s_c}, even in the altruistic game $s_c(\beta)$, where $\beta >> \alpha_{s_c}$, the social optima of $s_c(\beta)$ remains Nash. Under this formalism, we have a single, scalar, value Γ describing the selfishness level for the whole Markov game, taking a conservative view with respect to rating a Markov game's cooperativeness. If there is only a single state under which players are able to grossly exploit their peers then the selfishness level of the game becomes, principally, defined by that interaction alone.

4 Experiments

We present our empirical analysis studying the effect of a selfishness level inspired reward shaping mechanism in two well-known SeqSDs, 'cleanup' and 'harvest' [5] (*public goods* and *commons* dilemmas [7], respectively) with code adapted from [13].

In both cleanup and harvest, agents are tasked with collecting apples that lie in an orchard. For both scenarios, rewards are acquired exclusively through the collection of apples, with the respective dilemmas arising from the means through which the pool of available apples is replenished. Agents are also able to 'zap' each other. Being zapped causes players to both receive negative reward and, if zapped multiple times in succession, are removed from play for some time.

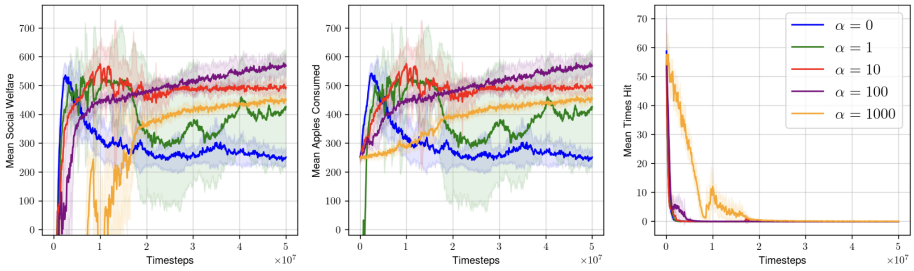


Fig. 1. Performance of varying α values under the harvest environment. Bold lines represent the rolling average of the respective metric over 5 runs with the shaded areas surrounding representing the standard deviation

4.1 Setup

We use proximal policy optimisation (PPO) [12] as the base learning algorithm for our policies with agents sharing network parameters. For each environment, we ran experiments with values for $\alpha \in \{0, 1, 10, 100, 1000\}$ as an exact deduction of Γ is infeasible. The primary metric used to judge the population of agents' tendency to cooperate is the social welfare (SW):

$$SW_t = \sum\nolimits_{i \in N} R_t^i$$

where, R_i is the extrinsic reward given by the environment to agent i at time t. We also plot: the number of apples consumed (AC), the number of times agents are hit with a zapper (Z), the Gini coefficient of apples consumed ($Gini$) and the amount of pollution cleaned (P):

$$AC_t = \sum\nolimits_{i \in N} a_t^i \qquad Z_t = \sum\nolimits_{i \in N} z_t^i$$

$$Gini_t = \frac{\sum_{i \in N} \sum_{j \in N} |\sum_T a_t^i - \sum_T a_t^j|}{2N \sum_i a_t^i} \qquad P_t = \sum\nolimits_{i \in N} p_t^i,$$

where, $a_t^i, z_t^i, p_t^i = 1$ if agent i has collected an apple, been zapped or removed a tile of pollution at time t respectively and 0 otherwise.

4.2 Results

Figure 1 shows our results in harvest. Agents with $\alpha > 0$ tend to outperform agents with $\alpha = 0$ but when $\alpha = 1$ or $\alpha = 10$, there is a large variance in performance between runs. $\alpha = 100$ results in the best performing agents at the end of training.

Figure 2 shows our results in cleanup. Here, we observe strikingly improved social welfare from agents with $\alpha = 100$. Even though there is no reward for doing so, our methodology successfully instils agents with an incentive to act in the public good. This is evidenced by the increased tendency to clean pollution when $\alpha > 0$, providing justification that our method increases the incentive for cooperative behaviour to emerge. We also find that, in both environments, agents

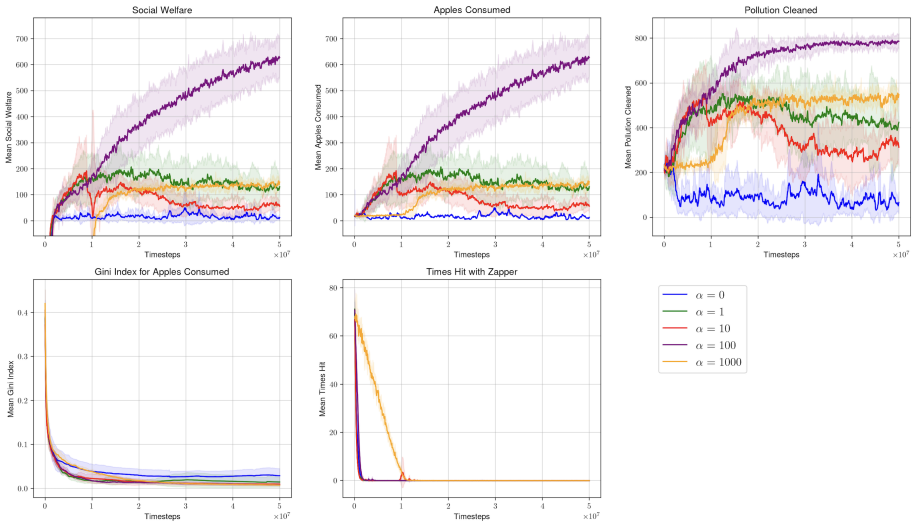


Fig. 2. Performance of varying α values under the cleanup environment. Bold lines represent the rolling average of the respective metric over 5 runs with the shaded areas surrounding representing the standard deviation

quickly learn to avoid zapping. In the case of $\alpha > 0$, this is likely due to the negative reward associated with being zapped whereas when $\alpha = 0$ the lack of positive feedback associated with the action likely causes zap-heavy policies to die-out in favour of apple-consuming ones. i.e. Zapping wastes time that could be better spent acquiring apples. This drop-off is not so quickly realised with $\alpha = 1000$ - we suggest this could be due to reward inflation muddying the distinction between good and bad policies. We initially assumed that social welfare would increase monotonically with α. We observe that agents learning under $\alpha = 1000$ perform strictly worse than those with $\alpha = 100$ and, in cleanup, agents with $\alpha = 10$ perform worse on average than those with $\alpha = 1$ suggesting that the value of α is indeed meaningful.

5 Conclusion

In this work, we explore the effectiveness of analysing SDs through the lens of their selfishness levels. We derive some interesting properties of SDs in the normal-form case, finding exact conditions under which selfishness-level-based payoff modifications can result in complete resolution of the dilemma. We further extend this work by providing a first-step towards a selfishness level in SeqSDs. Our empirical results suggest a strong benefit to the cooperative performance of learning agents in SeqSDs with the overall impact of our method being to add additional, socially optimal, equilibria to the policy space but not to prescribe any particular best solution. As such, we suspect an *equilibrium selection* problem is still present within our method.

References

1. Apt, K.R., Schäfer, G.: Selfishness level of strategic games. In: Serna, M. (ed.) Algorithmic Game Theory. pp. 13–24. Springer Berlin Heidelberg, Amsterdam, NL (2012)
2. Christoffersen, P.J., Haupt, A.A., Hadfield-Menell, D.: Get it in writing: Formal contracts mitigate social dilemmas in multi-agent rl. In: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems. p. 448–456. AAMAS '23, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2023)
3. Dawes, R.M.: Social dilemmas. Annual review of psychology **31**(1), 169–193 (1980)
4. H. Kelley, H., W. Thibaut, J.: Interpersonal Relations: A Theory of Interdependence. John Wiley & Sons, NY, United States (1978)
5. Hughes, E., Leibo, J.Z., Phillips, M., Tuyls, K., Dueñez Guzman, E., Castañeda, A.G., Dunning, I., Zhu, T., McKee, K., Koster, R., Roff, H., Graepel, T.: Inequity aversion improves cooperation in intertemporal social dilemmas. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. p. 3330–3340. NIPS'18, Curran Associates Inc., Red Hook, NY, USA (2018)
6. Kollock, P.: Social dilemmas: The anatomy of cooperation. Annual review of sociology **24**(1), 183–214 (1998)
7. Kollock, P.: Social dilemmas: The anatomy of cooperation. Annual Review of Sociology **24**(1), 183–214 (1998). https://doi.org/10.1146/annurev.soc.24.1.183
8. Leibo, J.Z., Zambaldi, V., Lanctot, M., Marecki, J., Graepel, T.: Multi-agent reinforcement learning in sequential social dilemmas. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. p. 464–473. AAMAS '17, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2017)
9. Macy, M.W., Flache, A.: Learning dynamics in social dilemmas. Proceedings of the National Academy of Sciences of the United States of America **99**(10), 7229–7236 (2002), `http://www.jstor.org/stable/3057846`
10. Madhushani, U., McKee, K.R., Agapiou, J.P., Leibo, J.Z., Everett, R., Anthony, T., Hughes, E., Tuyls, K., Duéñez-Guzmán, E.A.: Heterogeneous social value orientation leads to meaningful diversity in sequential social dilemmas. arXiv preprint arXiv:2305.00768 **0**, 9 (2023)
11. McKee, K.R., Gemp, I., McWilliams, B., Duèñez Guzmán, E.A., Hughes, E., Leibo, J.Z.: Social diversity and social preferences in mixed-motive reinforcement learning. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems. p. 869–877. AAMAS '20, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2020)
12. Schulman, J., Wolski, F., Dhariwal, P., Radford, A., Klimov, O.: Proximal policy optimization algorithms. CoRR **abs/1707.06347**, 12 (2017), `http://arxiv.org/abs/1707.06347`
13. [Vinitsky, E., Jaques, N., Leibo, J., Castenada, A., Hughes, E.: An open source implementation of sequential social dilemma games. `https://github.com/eugenevinitsky/sequential_social_dilemma_games/issues/182` (2019), gitHub repository
14. Vinitsky, E., Köster, R., Agapiou, J.P., Duéñez-Guzmán, E.A., Vezhnevets, A.S., Leibo, J.Z.: A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. Collective Intelligence **2**(2), 26339137231162025 (2023)

15. Wang, J.X., Hughes, E., Fernando, C., Czarnecki, W.M., Duéñez Guzmán, E.A., Leibo, J.Z.: Evolving intrinsic motivations for altruistic behavior. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. p. 683–692. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019)

A Proofs of Theorems

To simplify our following analysis we re-state the definition of social dilemma. We construct a new game, without loss of generality, by applying the positive affine transformation $p_i(s) - S,\ \forall s \in \{S_i\}_{i \in N}$:

	C	D
C	$R-S, R-S$	$S-S, T-S$
D	$T-S, S-S$	$P-S, P-S$

and simplify notation:

	C	D
C	R, R	$0, T$
D	$T, 0$	P, P

and finally, re-write the social dilemma inequalities as:

$$R > P,\ R > 0,\ 2R > T, \tag{4}$$

$$T > R \text{ (greed) or } P > 0 \text{ (fear)}. \tag{5}$$

A.1 The Proof of Theorem 1

The selfishness level of a SD is

$$\alpha_G = \begin{cases} 0 & \text{if } T \leq R, \\ \frac{T-R}{2R-T} & \text{if } T > R. \end{cases} \tag{6}$$

Proof. Recall that the unique, stable, social optimum of a social dilemma is obtained through mutual cooperation (Equation 4), easing this process to simply finding the exact α under which (C, C) becomes Nash. Also recall that there exist three, distinct, modalities of social dilemma:

1. $T > R$ and $P > S$: *Prisoner's Dilemmas*
2. $T > R$ and $P \leq S$: *Chicken Dilemmas* and,
3. $T \leq R$ and $P > S$: *Stag Hunt Dilemmas.*

It is straightforward to see that, in stag hunt dilemmas, (C, C) is a Nash equilibrium. This means that, if the social dilemma is a stag hunt, the selfishness level is $\alpha_G = 0$. To see the selfishness level in prisoner's and chicken dilemmas, we introduce notation for the payoffs of the altruistic modification of G, $G(\alpha)$ (see Table 3).

	C	D
C	R', R'	S', T'
D	T', S'	P', P'

Table 3. Payoff matrix for $G(\alpha)$

where,

$$R' = R + \alpha 2R,$$
$$T' = T + \alpha T,$$
$$S' = \alpha T,$$
$$P' = P + \alpha 2P.$$

For both prisoner's dilemmas and chicken dilemmas, (C, C) is not a Nash equilibrium in G as $T > R$. For (C, C) to be a Nash equilibrium in $G(\alpha)$, the following must hold:

$$\begin{aligned} R' &\geq T' \\ R' - T' &\geq 0 \\ (R + \alpha 2R) - (T + \alpha T) &\geq 0 \end{aligned} \tag{7}$$

Changing the inequality to an equality and solving for α gives us the lowest bound on α which satisfies the condition:

$$\alpha = \frac{T - R}{2R - T} \tag{8}$$

A.2 The Proof of Theorem 2

Given a SD G, let $T > R$ and $P \leq 0$ (a chicken dilemma). $G(\alpha)$ is always resolved when $\alpha = \alpha_G$.

Proof. Given a chicken dilemma, we have the following payoffs in $G(\alpha_G)$

$$R' = R + \frac{T - R}{2R - T} 2R, \tag{9}$$
$$T' = T + \frac{T - R}{2R - T} T, \tag{10}$$
$$S' = \frac{T - R}{2R - T} T, \tag{11}$$
$$P' = P + \frac{T - R}{2R - T} 2P. \tag{12}$$

For $G(\alpha_G)$ to be resolved, we need to have $T' \leq R'$ and $P' \leq S'$. Given equations 7 and 8, we already have that $T' = R'$. The second inequality follows from the following claims

- Claim 1: $S' > 0$.Recall that in a chicken dilemma $T > R$. As $T > R > 0$ and $2R > T$, $\frac{T-R}{2R-T} > 0$ hence $S' > 0$.
- Claim 2: $P' \leq 0$. Recall that in a chicken dilemma, $P \leq 0$. If $P = 0$, $P' = 0$. If $P < 0$, $P' < 0$

Combining the above claims, we get that $P' < S'$. Hence, $G(\alpha_G)$ is resolved.

A.3 The Proof of Theorem 3

Given a SD G, let $T > R$ and $P > 0$ (a prisoner's dilemma). $G(\alpha)$ is resolved when $\alpha = \alpha_G$ and $P \leq T - R$

Proof. Given a prisoner's dilemma, we have payoffs consistent with equations 9 - 12 in $G(\alpha_G)$. For $G(\alpha_G)$ to be resolved, $T' \leq R'$ and $P' \leq S'$. Given equations 7 and 8, $T' = R'$. We can set $P' \leq S'$ and simplify to find the appropriate bound:

$$P' \leq S' \implies P + \frac{T-R}{2R-T}2P \leq \frac{T-R}{2R-T}T \tag{13}$$

which after some simple algebra leads to $P \leq T - R$ as claimed.

3.2 Fostering Multi-Agent Cooperation through Implicit Responsibility

Fostering Multi-Agent Cooperation through Implicit Responsibility

Daniel E. Collins[0000−0002−1075−4063], Conor Houghton[0000−0001−5017−9473], and Nirav Ajmeri[0000−0003−3627−097X]

University of Bristol, Bristol, UK
{daniel.collins,conor.houghton,nirav.ajmeri}@bristol.ac.uk

Abstract. For integration in real-world environments, it is critical that autonomous agents are capable of behaving responsibly while working alongside humans and other agents. Existing frameworks of responsibility for multi-agent systems typically model responsibilities in terms of adherence to explicit standards. Such frameworks do not reflect the often unstated, or implicit, way in which responsibilities can operate in the real world. We introduce the notion of *implicit responsibilities*: self-imposed standards of responsible behaviour that emerge and guide individual decision-making without any formal or explicit agreement.
We propose that incorporating *implicit responsibilities* into multi-agent learning and decision-making is a novel approach for fostering mutually beneficial cooperative behaviours. As a preliminary investigation, we present a proof-of-concept approach for integrating *implicit responsibility* into independent reinforcement learning agents through reward shaping. We evaluate our approach through simulation experiments in an environment characterised by conflicting individual and group incentives. Our findings suggest that societies of agents modelling *implicit responsibilities* can learn to cooperate more quickly, and achieve greater returns compared to baseline.

1 Introduction

When tasked with navigating complex social decision-making scenarios alongside humans and other agents, it is important that agents can balance potential incentive conflicts, and find ways to perform their allocated role effectively whilst acting in a manner that is considered responsible and ethical by human standards [4, 11]. Existing works have outlined various facets of responsibility in multi-agent systems (MAS) [12].

Responsibility. A general definition of responsibility, outlined in [12], involves the expectation for an agent or group of agents, A, to realise a future state, φ, of the environment [5, 8].

Explicit Responsibility. Typically, responsibilities are modelled in terms of standards of behaviour that are prescribed "top-down", such as accountability for the fulfilment of allocated tasks or sanctionability for the violation of a social norm [12]. In this paradigm, agents are responsible to the extent that they adhere to an explicit system

of rules. Similarly, responsibility can be imposed through explicit agreements or commitments between agents [1, 6]. We group these treatments as *explicit responsibility*, which can always be described by "A is responsible for φ under z", where z represents the explicit source of the responsibility, which may be enforced top-down, agreed upon peer-to-peer, or otherwise entered into knowingly.

Example 1 (Explicit Responsibility). Alice adopts a puppy in the UK. By adopting the puppy, Alice has agreed to an explicit duty of care; they are aware that they are accountable for the welfare of the dog under UK law, and that adopting and subsequently neglecting a dog would violate social convention. If Alice proceeds to neglect the puppy, they may be subject to legal repercussions, or disapproval and alienation from family and friends.

Implicit Responsibility. In contrast to *explicit responsibility*, relatively little attention has been given to aspects of responsibility that emerge without any imposed standards or explicit agreement between parties. Self-imposed responsibilities can play an important role in ethical decision making amongst people. Affective responses to different scenarios and outcomes can reinforce an individual sense of responsibility, motivating subsequent cooperation and altruistic behaviour. Individual differences in these affective responses can give rise to variations in self-motivated responsible behavior between people. Understanding this type of responsibility and how it can lead to alignment and misalignment of individual perceptions of responsibility in society is important for citizen-centric design of MAS. We extend the conceptual framework of *explicit responsibility* in MAS by introducing the notion of *implicit responsibility*: a self-imposed responsibility for bringing about some φ, that emerges bottom-up, and is internally motivated and voluntarily assumed without any explicit mandate, commitment or expectation.

Example 2 (Implicit Responsibility). Alice comes across a stunned pigeon near their home. Alice reasons that the pigeon will likely be in danger if left in its current state, and that they could carefully transfer the pigeon to a cardboard box and leave it to rest in a safe quiet area to recover. Alice is driven to help the pigeon by an internal sense of responsibility, although there is no explicit expectation to do so.

In Example 2, a situation emerges in which Alice feels implicitly responsible for the fate of another entity. Even if Alice does not assume the responsibility for assisting the other entity as a goal, they are nevertheless aware that they are capable of providing that assistance, and the consequences of not doing so. Failure to help may confer a negative affective state, motivating Alice to help in similar scenarios in the future.

Contributions. In this work, we introduce the notion of *implicit responsibility* in MAS. We present a novel approach for promoting cooperation within the framework of multi-agent reinforcement learning (MARL) by operationalising *implicit responsibility* for reward shaping. We investigate our approach by conducting simulation experiments in a constrained task environment designed to incorporate well-defined *implicit responsibilities*. We compare the learning of cooperative behaviour by *implicit responsibility* agents to *baseline* reinforcement learning agents that do not shape rewards. We find

that agents that model *implicit responsibility* learnt cooperative strategies faster, and demonstrate improved performance on the task compared to *baseline* agents.

2 Operationalising Implicit Responsibility in MAS

In MARL, reward shaping is the process of modifying an agent's reward function by introducing additional "pseudo-rewards" to guide agents towards learning specific patterns of behaviour that may not be adequately incentivised by the original reward function. Shaping rewards according to violation or satisfaction of *implicit responsibility* provides a novel framework for learning desirable behaviour. For a pair of agents A, B, A has an *implicit responsibility*, $R^t_{A,B}(\varphi_B)$, for realising a future state of the environment, φ_B, if at some time, t, the environment state, s^t satisfies all of three conditions

1. *Existence of Dependency*, $\psi_{A,B}(1)$ - Agent B's ability to achieve their goals in a future state $s \in \varphi_B$ is contingent on the actions or resources of A.
2. *Capability to Influence*, $\psi_{A,B}(2)$ - Agent A possesses the capacity to address the needs of B and bring about φ_B through its actions or resources.
3. *Awareness or Capability of Perception*, $\psi_{A,B}(3)$ - Agent A can perceive or is capable of perceiving conditions (1) and (2) even if B does not communicate this explicitly.

These conditions describe circumstances in which the realisation of some φ_B, in which B can pursue their goals without assistance, is not possible through the actions of B alone, or from the influence of the dynamics of the environment itself.

2.1 Foraging Survival Simulation Environment

We designed a multi-agent grid-world environment that incorporates well-defined opportunities for *implicit responsibilities*, as an evaluation test-bed. The environment is illustrated in Figure 1.

Setup In this environment, a population of agents, I, navigate an M by N grid-world with the goal of collecting berries. Initially, each agent $i \in I$ starts from a random empty position, and $|I|$ berries are placed at random empty positions so that the number of berries is equal to the number of agents.

Agent attributes Agents have two attributes which relate to their survival in the environment: (1) energy and (2) health. These are represented by the integers $e_i \in \mathbb{Z}^+ : e_i \in [0, E]$ and $h_i \in \mathbb{Z}^+ : h_i \in [0, H]$ respectively. Agents are initialised with $e_i = E$ and $h_i = H$.

Attribute decay Agents are in one of three possible states at any time, based on their attributes: (1) *Healthy*: $(e_i > 0, h_i = H)$, (2) *Helpless*: $(e_i = 0, h_i > 0)$, and (3) *Dead*: $(e_i = 0, h_i = 0)$. While agents are *Healthy*, e_i decays by one per time step. When $e_i = 0$, agents become *Helpless*, and h_i begins to decay by one per time step. Agents can only take actions while *Healthy*. If the agent transitions into the *Dead* state, $h_i = 0$, the agent is removed from the simulation for the remainder of the episode.

Berry collection Agents collect berries by moving to their positions. When an agent collects a berry, the agent receives a reward r_b, and a new berry is generated at a random

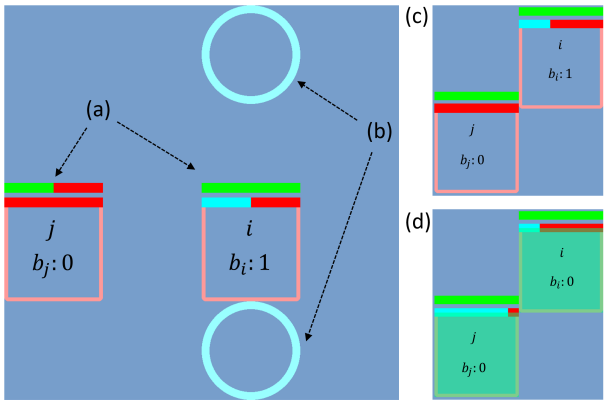


Fig. 1: (Left) Two agents i and j (a) navigate a 4×4 grid world and collect berries (b). The agents health h_i, h_j and energy e_i, e_j are indicated by indicated by the upper and lower bars above the agents respectively, and the number of stored berries is indicated by b_i, bj. (Right) (c) In the illustrated scenario, j has $e_j = 0$, and no stored berries, and i has $e_i > 0$ and one stored berry. (d) In the next time step, i throws their stored berry to j, illustrated by the green shading, and e_j is restored.

unoccupied position. If an agent dies, the next berry collection will not trigger a new berry to be generated. This ensures that there is only one berry per living agent in the environment.

Berry inventory Agents *store* collected berries in an inventory. The number of stored berries is $b_i \in \mathbb{Z}^+ : b_i \in [0, B]$, where B is the inventory capacity.

Berry consumption Agents *consume* stored berries to fully restore e_i and h_i. If an agent has $b_i > 0$ when $e_i = 0$, the agent automatically *consumes* a stored berry. Agents therefore have an *effective energy* of $e'_i = e_i + E * b_i$.

Agent actions Agents have five discrete movement actions for navigating the environment:*up*, *down*, *left*, *right*, and *stay*. Additionally, agents have a *throw* action which passes a stored berry to the agent, j, with the lowest *effective energy*, e'_j. If $b_i = 0$, or if all other agents are *dead*, the *throw* fails and the berry remains in the agents inventory. If an agent successfully throws a berry, their energy does not decay in that time step.

Decision module Agents automatically *consume* a berry if: (1) $h_i < H$ and $b_i > 0$ at the start of a time step, (2) $h_i < H$ and i has just been passed a berry by another agent, or (3) $b_i = B$ and i has just collected a new berry.

Agents have an immediate incentive to act in self-interest by collecting berries as quickly as possible. However, the *Throw* mechanic allows *Healthy* agents to cooperate by paying a cost to revive *Helpless* agents and prevent their death. We can introduce a long-term incentive for mutual cooperation which outweighs the immediate incentive for self-interest through careful choice of environment parameters, (M, N), E, H, B and $|I|$. In Appendix B, we choose environment parameters for our experiment such that mutual cooperation can facilitate longer survival times, and thus greater overall returns.

2.2 Reward Shaping using Implicit Responsibility Conditions

We now apply the conditions described in Section 2 for formation of *implicit responsibility* to our environment. For two agents $i, j \in I$, let φ_j be the set of states in which j is *Healthy*, such that $s_j^t \in \varphi_j$ if $h_j^t = H$. $R_{i,j}^t(\varphi_j) = R_{i,j}^t$ then describes whether i has an *implicit responsibility* towards j at time t for realising φ_j if all three conditions (*Existence of Dependency*, *Capability to Influence*, *Awareness or Capability of Perception*) are met.

For our environment, the condition $\psi_{i,j}^t(1)$ for *Existence of Dependency* is true if j has no energy or berries, but is not yet *Dead*.

$$\psi_{i,j}^t(1) = \begin{cases} 1, & \text{if } e_j^t = 0, \text{ AND } b_j^t = 0, \text{ AND } h_j^t > 0 \\ 0, & \text{otherwise} \end{cases}$$

The condition $\psi_{i,j}^t(2)$ for *Capability to Influence* is true if i has enough energy and berries to throw one to j, and i will not run out of energy as a result of the throw. Let ω_i^t be the *Spare Effective Energy* of i at t, e.g. the effective energy of i that would remain after throwing a berry, $\omega_i^t = e_i^t + E \cdot (b_i^t - 1)$. Let k_i^t be the shortest Manhattan distance between i and any berry at t. If $k_i^t < \omega_i^t$, i can throw a berry and have enough energy remaining to reach another.

$$\psi_{i,j}^t(2) = \begin{cases} 1, & \text{if } k_i^t < \omega_i^t \\ 0, & \text{otherwise} \end{cases}$$

For $\psi_{i,j}(3)$, *Awareness or Capability of Perception*, we assume full-observability of the environment for all agents, therefore i always has sufficient information to know if $\psi_{i,j}(1)$ and $\psi_{i,j}(2)$ are true, thus $\psi_{i,j}(3)$ is true by default.

Once formed, an *implicit responsibility* is maintained until the next time step in which any of the individual conditions are broken. If a responsibility is formed at a time t and maintained until any condition is broken at some later time t', the responsibility is violated if the state $s^{t'}$ does not belong to φ_j. Otherwise, if $s^{t'} \in \varphi_j$, the responsibility is satisfied. Algorithm 1 in Appendix A describes our method for shaping rewards by applying penalties, p, for violating an *implicit responsibility*.

3 Simulation Experiments

We conduct preliminary simulation experiments using the environment described in Section 2.1 with the parameters outlined in Appendix B, Table 1. We simulate and compare societies comprising pairs of agents, which are trained using Deep Q-Learning as described in Appendix C, with hyper-parameters in Table 2. We train a *baseline* agent society using only extrinsic rewards signals from berry collection, and an *implicit responsibility* agent society using both extrinsic rewards and additional penalties for violation of *implicit responsibilities* using our reward shaping algorithm (Section A, Algorithm 1). To evaluate our *implicit responsibility* agents, we compare the length of each episode during training to those achieved by *baseline* agents. Episode length tell us the total survival time of an agent society, indicating the performance of the agents during training.

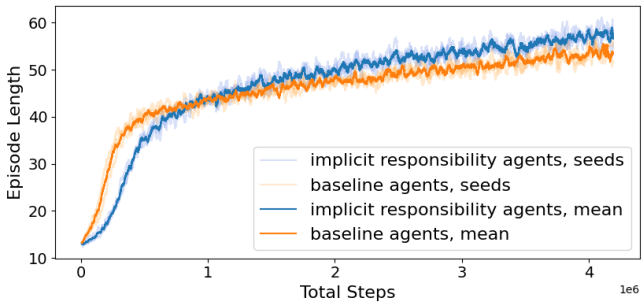


Fig. 2: Episode length (moving average, window size = 1000) vs total environment steps elapsed during training. The mean across three random seeds is shown alongside each individual seed.

4 Discussion

Figure 2 shows the training curves for *baseline* agents and *implicit responsibility* agents across three random seeds. For each episode during training, the episode length is plotted against the the total number of time steps that have elapsed prior to the episode during training. In the early stages of training, *baseline* agents achieve greater survival times than *implicit responsibility* agents. However, after roughly 10^6 steps, *implicit responsibility* agents demonstrate greater survival times on average. These results are a promising indication that shaping rewards according to *implicit responsibility* can improve the speed at which reinforcement learning agents learn to exploit mutually beneficial cooperation behaviours. However, there are several limitations which must be addressed. Firstly, we only evaluate under one set of environment parameters and learning hyper-parameters. It is possible that the benefits of our approach are less significant when we compare to baseline under an optimised training protocol, or in societies of more than two agents. Further experimentation would be needed to validate our findings and assess scalability.

Further, we only test in one environment, which we designed to include easily defined scenarios for *implicit responsibility* to arise, and in which cooperation is globally beneficial. In doing so, we were able to test our approach by shaping rewards according to rules representing an idealised and thus *explicit* model of *implicit responsibility* for that environment. For application to unseen and more complex environments, agents must be designed such that they are able to approximate these rules independently. Causal attribution of responsibility and blameworthiness for outcomes are non-trivial problems [9, 12], posing a challenge for reward function design.

Finally, we consider only a subset of *implicit responsibilities* that capture mutually beneficial outcomes, and thus neglects the role of altruism captured by other approaches for bottom-up learning of responsible behaviour [2, 3, 10].

Bibliography

[1] Dastani, M., van der Torre, L., Yorke-Smith, N.: Commitments and interaction norms in organisations. Autonomous Agents and Multi-Agent Systems (JAAMAS) **31**(2), 207–249 (Mar 2017)

[2] Deshmukh, J.: Emergent responsible autonomy in multi-agent systems. In: Proc. AAMAS. pp. 3029—3031. (May 2023)

[3] Deshmukh, J., Adivi, N., Srinivasa, S.: Resolving the dilemma of responsibility in multi-agent flow networks. In: Proc. PAAMS. pp. 76–87. (Jul 2023)

[4] Murukannaiah, P.K., Ajmeri, N., Jonker, C.M., Singh, M.P.: New foundations of ethical multiagent systems. In: Proc. AAMAS. pp. 1706–1710. (May 2020)

[5] van de Poel, I.: The Relation Between Forward-Looking and Backward-Looking Responsibility. In: Moral Responsibility: Beyond Free Will and Determinism, pp. 37–52. Library of Ethics and Applied Philosophy, Springer (2011)

[6] Singh, M.P.: Norms as a basis for governing sociotechnical systems. ACM Transactions on Intelligent Systems and Technology (TIST) **5**(1), 21:1–21:23 (Dec 2013)

[7] Szita, I., Lőrincz, A.: The many faces of optimism: a unifying approach. In: Proc. ICML. pp. 1048–1055. ACM (2008)

[8] Triantafyllou, S.: Forward-Looking and Backward-Looking Responsibility Attribution in Multi-Agent Sequential Decision Making. In: Proc. AAMAS. pp. 2952–2954 (May 2023)

[9] Triantafyllou, S., Radanovic, G.: Towards Computationally Efficient Responsibility Attribution in Decentralized Partially Observable MDPs. In: Proc. AAMAS. pp. 131–139. (May 2023)

[10] Wang, J.X., Hughes, E., Fernando, C., Czarnecki, W.M., Duenez-Guzman, E.A., Leibo, J.Z.: Evolving intrinsic motivations for altruistic behavior (Mar 2019), arXiv:1811.05931 [cs]

[11] Woodgate, J., Ajmeri, N.: Macro ethics for governing equitable sociotechnical systems. In: Proc. AAMAS. pp. 1824–1828. (May 2022).

[12] Yazdanpanah, V., Gerding, E.H., Stein, S., Cirstea, C., Schraefel, M.C., Norman, T.J., Jennings, N.R.: Different Forms of Responsibility in Multiagent Systems: Sociotechnical Characteristics and Requirements. IEEE Internet Computing **25**(6), 15–22 (Nov 2021)

[13] Zhu, Z., Hu, C., Zhu, C., Zhu, Y., Sheng, Y.: An Improved Dueling Deep Double-Q Network Based on Prioritized Experience Replay for Path Planning of Unmanned Surface Vehicles. Journal of Marine Science and Engineering **9**(11), 1267 (Nov 2021)

A Reward Shaping Algorithm

Algorithm 1 Reward shaping for *implicit responsibility* agents

Let i, j be any pair of agents from a population I.
Let b_i^t be the number of berries that i has stored in their inventory at time t, where $0 \leq b_i^t \leq B$ and $b_i^t, B \in \mathbb{Z}^+$
Let e_i^t be the energy of i at t, where $0 \leq e_i^t \leq E$ and $e_i^t, E \in \mathbb{Z}^+$
Let h_i^t be the health of i at t, where $0 \leq h_i^t \leq H$ and $h_i^t, H \in \mathbb{Z}^+$
Let $d_{i,j}^t$ be the Manhattan distance between i and j at t.
Let k_i^t be the shortest Manhattan distance between i and any berry at t.
Let ω_i^t be the *Spare Effective Energy* of i at t, where $\omega_i^t = e_i^t + E \cdot (b_i^t - 1)$
Let s^t represent the full environment state at time t.
Let r_i^t be the reward to i at time t.
Let p be the constant representing the penalty for violation of an *implicit responsibility*.
Let φ_j be the set of states in which j is *Independent*, such that $s_j^t \in \varphi_j$ if $h_j^t = H$.
Let $\psi_{i,j}^t(1)$ describe the condition for the *Existence of Dependency* such that

$$\psi_{i,j}^t(1) = \begin{cases} 1, & \text{if } e_j^t = 0, \text{ AND } b_j^t = 0, \text{ AND } h_j^t > 0 \\ 0, & \text{otherwise} \end{cases}$$

Let $\psi_{i,j}^t(2)$ describe the condition for *Capability to Influence* such that

$$\psi_{i,j}^t(2) = \begin{cases} 0, & \text{if } k_i^t > \omega_i^t \\ 1, & \text{otherwise} \end{cases}$$

Let $R_{i,j}^t$ be the bool representing whether i has an *implicit responsibility* towards j at time t

$$R_{i,j}^t = \begin{cases} True, & \text{if } \psi_{i,j}^t(1) = 1, \text{ AND } \psi_{i,j}^t(2) = 1 \\ False, & \text{otherwise} \end{cases}$$

// Iterate over all permutations of agent pairs $i, j \in I$
for $i \in I$ **do**
  for $j \in I : j \neq i$ **do**
    // If i was responsible before but not after the transition ...
    if $R_{i,j}^t$ AND $\neg \mathrm{R}_{i,j}^{t+1}$ **then**
      //...and if j has not reached φ_j
      if $\neg(\mathrm{s}^{t+1} \in \varphi_j)$ **then**
        // Apply penalty for violation
        $r_i^{t+1} = r_i^{t+1} - p$

B Environment Parameters

For an (M, N) grid with population $|I|$, if we do not allow agents to use the *Throw* action, and if E is less than some threshold, E^*, the energy of each agent will on average decay towards zero each time step, and all agents will eventually die even with

an optimal coordinated foraging strategy. For our environment, we estimate E^* to be the average Manhattan distance between any agent and their closest berry for all possible combinations of positions of $|i|$ agents and $|I|$ berries. In practice, E^* will be slightly lower since the optimal foraging strategy would also ensure that no two or more agents target the same berry at any time. By allowing agents to *Throw* berries, the population can cooperate to survive for longer and thus achieve greater overall returns. For our experiments, we use the environment parameters shown in Table 1.

Table 1: Default environment parameters.

Parameter	**Default Value**
Grid Shape (M, N)	(4, 4)
Population Size $\|I\|$	2
Max Energy E	2
Max Health H	6
Inventory Capacity B	10
Berry Reward r_b	0.1
Violation Penalty p	-0.9

C Agent Architecture and Hyperparameters

Here we describe a schematic of the modular architecture used for our *baseline* and *implicit responsibility* agents. In our experiments, both *baseline* and *implicit responsibility* agents are trained using using independent Deep Q learning implemented with PyTorch. Agents comprise a Deep Q-Network (DQN) architecture with with two fully connected layers. We employ experience replay [13] to stabilise the learning process. Agents explore their shared environment using an epsilon-greedy [7] exploration strategy with exponential decay. Table 2 lists the hyper-parameters of the learning procedure.

Table 2: DQN hyperparameters.

Hyperparameter	**Value**
Batch Size	64
Replay Buffer Capacity	10 000
Discount Factor	0.99
Initial Exploration Rate	0.9
Final Exploration Rate	0.005
Exploration Steps	1000
Tau	0.005
Learning Rate	0.001
Loss Function	MSE
Target Network Update Frequency	500

4 Agent-Based Models and Human-Agent Interaction

4.1 Mitigating School Segregation through Targeted School Relocation

Mitigating School Segregation through Targeted School Relocation

Mayesha Tasnim[0000−0002−0127−4797], Dimitris Michailidis[0000−0002−0106−1126], Sennay Ghebreab[0009−0007−5788−4635], Fernando P. Santos[0000−0002−2310−6444], and Erman Acar[0000−0001−7541−2999]

Socially Intelligent Artificial Systems, University of Amsterdam
{m.tasnim, d.michailidis, s.ghebreab, f.p.santos, e.acar}@uva.nl

Abstract. School segregation tends to mimic and often surpass residential segregation. This trend persists even in cities with an open school choice policy. As school segregation exacerbates patterns of social inequality, policymakers need interventions that can foster positive behavioral changes to counter segregation trends. Previous work has suggested interventions based on housing support, zoning policies, and public transportation design. These interventions might however face challenges associated with housing shortages, slow implementation of zoning reforms, and budget constraints to augment public transportation. In this work, we propose a new type of intervention in the form of *targeted school relocation*. This approach involves relocating under-subscribed schools to specific nodes within the network to mitigate segregation. Preliminary result suggests that strategically relocating schools can reduce segregation more efficiently than transport network interventions.

Keywords: School Segregation · Algorithmic Interventions · Agent-based Simulations.

1 Introduction

The widespread adoption of open enrollment policies by public schools grants students the freedom to choose any school within a city. This coincides with many schools instituting internal integration policies aimed at cultivating diversity among student populations [10, 9]. School choice is influenced by various factors, such as the proximity of schools and perceived educational quality [4]. Notably, the phenomenon of homophily has emerged in recent years as a significant determinant shaping parental preferences for schools [17, 4].

As societies become progressively more diverse and multicultural, parental school choice has undergone a notable shift. Alongside traditional considerations like academic reputation and proximity, parents today increasingly prioritize schools based on peers of similar cultural or socioeconomic backgrounds [14, 2]. This trend towards homophilic school selection exacerbates existing patterns of racial and social inequality, increasing disparities in access to resources and opportunities for minority groups [6, 11, 14].

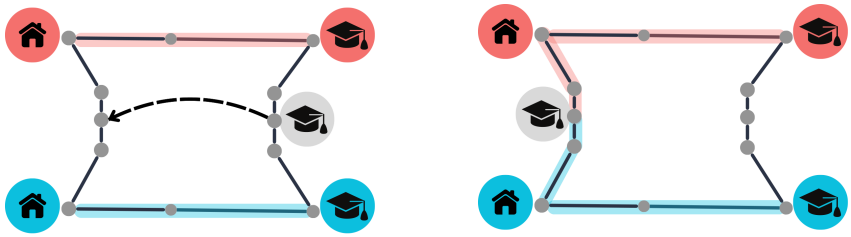


Fig. 1: **Targeted School Relocation**. A transportation network is depicted with nodes corresponding to neighborhoods containing residences or schools. The targeted school relocation algorithm identifies an under-subscribed school (marked in gray) and moves it between two segregated residential neighborhoods (marked in red and blue). This creates an educational alternative for students living in segregated neighborhoods, leading to increased heterogeneity in schools.

For school administrators, these societal dynamics present multifaceted challenges. Policymakers grapple with managing both under-subscribed and over-subscribed schools, having to decide strategically which schools to close, expand, merge, or relocate [12]. These interventions aim not only to mitigate financial risks in budgeting but also to foster greater diversity and inclusivity in education.

To this end, policymakers are increasingly turning to agent-based modelling to tackle these intertwined social challenges involving complex systems. These models simulate the intricate interactions between residential patterns, school choice behaviors, and racial segregation dynamics, offering valuable insights into the potential impacts of various policy interventions on urban education systems [3, 5, 16]. Recent works have explored interventions to reduce school segregation through housing support [8] and changing zoning policy [19]. School segregation can be reduced in specific contexts by changing individuals' preferences for homophilic school composition [15] or adding new connections to the public transportation network [13]. However, given the complexity of school segregation, these interventions might be unproductive in specific contexts: cities might already suffer from housing shortage problems, zoning reforms can be slow to implement, citizens' homophilic preferences can prevail and transport network interventions may face budget and geographic constraints in creating new lines. It is desirable therefore to explore alternative intervention mechanisms – complementary to the ones listed above – to reduce school segregation.

In this work-in-progress paper, we propose *Targeted School Relocation* as a novel approach to mitigate school segregation without modifying individuals' preferences, the structure of a public transportation network, or zoning policies. We propose an algorithmic method to implement Targeted School Relocation in a spatial graph encoding travel time between neighborhoods and schools' locations. We frame reducing school segregation as a facility location problem and strategically relocate under-subscribed schools so that they are available as an alternative to the most segregated schools in the city, as illustrated in Figure 1.

In addition to creating an alternative to segregated schools, relocating under-subscribed schools allows for the efficient usage of a limited education budget. We conduct experiments in a synthetic environment to show that targeted school relocation can lead to a significant reduction in segregation over time, compared to previously proposed network augmentations [13].

2 Methods

2.1 Agent-Based Model

We introduce an algorithm designed for targeted school relocation within an existing public transportation network. Our approach builds upon the Agent-based Model outlined in [13], which defines a graph-based environment for agents and schools with explicitly defined models for school choice and for measuring segregation.

Environment Let $\mathbb{G} = (V, E)$ represent the environment defined as an undirected graph. $V = \{v_1, ..., v_{n_{|v|}}\}$ denotes nodes representing neighbourhoods and $E = \{e_{i,j} | i, j \in V, i \neq j\}$ are edges representing travel connections. $A = \{a_1, ..., a_N\}$ represents a set of N agents, where each agent resides in node $v_a \in V$ and belongs to a socioeconomic group $g \in G$. For the sake of simplicity, this work assumes two socioeconomic groups $G = \{g_1, g_2\}$. Schools are denoted by $f \in F$ which are located in nodes $v_f \in V$.

School Choice Model At every round, each agent $a_i \in A$ creates a preference list $P_i \subseteq F$, over schools. The preference list is based on a utility function $U_{if}, f \in F$, and schools are sorted in descending order. Let $C : c_{g,f} \rightarrow \mathbb{R}$ define the composition of a school, denoting the fraction of agents from group g attending school f. We use the well-known Cobb-Douglas utility function, based on a function of school composition $C : c_{g,f} \rightarrow \mathbb{R}$ and travel time $t_{i,f}$ from the agent's residence to the school f. [6, 18]

$$U_{i,f} = c_{g,f}^{\alpha} \, t_{i,f}^{(1-\alpha)}, \tag{1}$$

where g denotes the group that agent a_i belongs to and $0 \leq \alpha \leq 1$ is a parameter that controls the weight of the group composition over the travel time $t_{i,f}$.

The preference lists for all agents are provided as input to an allocation method R. R is defined as a function $R : P \rightarrow (F \times A)$ which takes as input preferences $p \in P$ and provides a school assignment $Q \in (F \times A)$ that records which students are assigned to which schools. Random Serial Dictatorship (RSD) is a popular mechanism for matching between schools and students [1]. For our simulations we implement RSD and perform allocations at every round; schools have the overall capacity to allocate all students.

Algorithm 1 Targeted School Relocation

Input $\mathbb{G} = (V, E)$, $Q \in (F \times A)$
Output $\mathbb{G}' = (V \cup v', E)$

for $b = 1, 2, \ldots B$ **do**
  $f' \leftarrow argmin_f\{\, a \in A, (a, f) \in Q\}$ ▷ find school with least students
  for $g \in g_1, g_2$ **do**
    $f_g \leftarrow argmax_f\{\, a \in g, (a, f) \in Q\}$ ▷ find most segregated school
    $v_g \leftarrow \{v \mid a \in v, a \in f_g\}$ ▷ find residence node of students attending f_g
  end for
  $v' \leftarrow \{v \mid d(v', v_{g_1}) = d(v', v_{g_2})\}$ ▷ place f' equidistant to f_{g_1} and f_{g_2}
  assign f' to v'
end for

Measuring Segregation After each simulation round, schools are evaluated on segregation. To measure segregation, we use the Dissimilarity Index (DI), a measure that captures the differences in the proportions of agents from two groups assigned to a school [7]. DI is defined as follows:

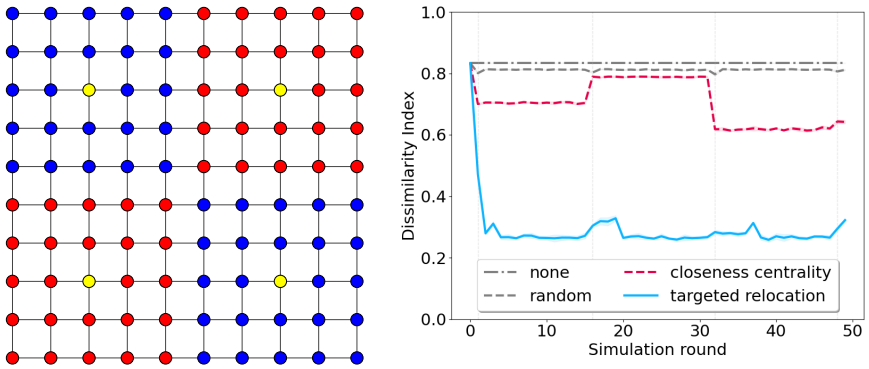
$$DI = \frac{1}{2} \sum_{f=1}^{|F|} \left| \frac{g_{1,f}}{G_1} - \frac{g_{2,f}}{G_2} \right|, \quad DI \in [0, 1] \tag{2}$$

where $g_{j,f}$ is the number of agents of group j in school f; G_j is the number of agents in group j. Segregation is minimum when $DI = 0$ and maximum when $DI = 1$.

2.2 Intervention Model

Transport Network Interventions We use the previously proposed transport interventions as a baseline to compare against our proposed model. These interventions were performed in the form of graph augmentations, by creating a new set of edges E' to be added to $\mathbb{G}$. The size of E' is equal to a budget $B \in \mathbb{N}$, which controls the number of allowed interventions per intervention round [13]. These interventions were considered a proxy for the creation of public transportation lines. A greedy algorithm was used to approximate the optimal set of interventions to apply to the graph with respect to accessibility. This translated to increasing a school node centrality $\mathbb{C}$ with respect to the other nodes. Two classes of greedy interventions, i.e., *centrality* and *group-based centrality optimization* were tested, and it was found that interventions based on *closeness centrality* performed best for reducing overall segregation [13].

Targeted School Relocation We introduce a new intervention method, which looks at relocating schools to existing nodes instead of creating new edges in $\mathbb{G}$. We refer to this as the **targeted school relocation** algorithm.



(a) Grid Environment (b) Simulation Results

Fig. 2: Simulation environment and results. The majority population in nodes are denoted by red and blue, with schools marked in yellow. Targeted school relocation can effectively decrease segregation, with fewer simulation rounds needed than interventions creating new edges using closeness centrality.

Let $F_g \subseteq F$ be a set of highly segregated schools such that each has a majority of students belonging to socioeconomic group $g \in G$. Conversely, let $F' \subseteq F \setminus F_g$ denote schools that receive the minimum number of students overall. Note that F' does not contain the segregated schools; instead, it is the set of under-subscribed schools that will be relocated to reduce the segregation in F_g. Let $d(v_1, v_2)$ denote the distance between two nodes.

The targeted school relocation algorithm identifies the residential nodes v_g of students attending F_g, and relocates F' to a node v' such that $d(v', v_{g_1}) = d(v', v_{g_2})$. This creates new options accessible to students from both neighborhoods, increasing the possibility they will be attended by students instead of F_g. Using under-subscribed schools F' also reduces the negative effect of relocating a school, such as displaced students. The procedure is described in Algorithm 1.

3 Experimental Setup

We implement our algorithm using the Agent-Based Model (ABM) outlined in Section 2 and a grid world environment, as shown in Figure 2a. The grid world environment is characterized by a lattice structure, with edges existing between all adjacent nodes. The environment is divided into two distinct communities: one occupying the northwest and southeast nodes and the other occupying the southwest and northeast nodes of the grid. One community is a 60% majority among the total population of 5000 agents. A high-residential segregation is induced in the environment by setting the majority population of a group in its respective node to 80% and the minority to 20% respectively. There are four

schools, each located at the centre of its respective community. The initial group composition of each school is set to be equal to the group composition of the node it is located in.

The ABM simulates the choice of schools by agents representing households and tracks the evolution of segregation patterns over multiple iterations. We compare the performance of our algorithm against baseline scenarios without school relocation, as well as alternative intervention strategies such as random school relocation.

4 Preliminary Results

Our initial experiments suggest that targeted school relocation can significantly reduce school segregation compared to baseline scenarios. Despite not introducing new transport network edges, the strategic placement of schools along key transit corridors effectively expands educational opportunities for students from segregated neighborhoods. In Figure 2b, we present the progress of the Dissimilarity Index (DI) over 50 simulation rounds. In each round, we perform 5 allocation rounds and report the 95% confidence interval.

Under the settings outlined in Section 3, the targeted school relocation algorithm leads to a significant reduction of DI, compared to a no-intervention scenario (none), the random relocation of a school (random), and the baseline intervention method of creating new transport network edges through greedily optimizing the closeness centrality of a school. It is also observed that targeted school relocation achieves a lower DI in fewer simulation rounds, suggesting that in certain conditions it can be more effective in reducing segregation than previously explored interventions.

The robustness of this method needs to be tested against more complex graph environments, based on real-world data. The regularity and connectivity of a grid environment facilitate the ideal placement of a school. This method, however, might face limitations when applied to heterogeneous spatial networks where nodes (i.e., neighbourhoods) might be characterized by different degrees of connectivity and centrality. Additionally, the impact of displaced students due to relocating under-subscribed schools is also not tested.

5 Conclusion and Future Work

In this work-in-progress paper, we present targeted school relocation as an approach to mitigate school segregation within existing public transportation networks, posing it as a cost-effective alternative to other forms of interventions, such as modifying school zones and creating new transport network lines. Our preliminary results using agent-based simulations suggest that this approach shows promising improvements over previously proposed interventions.

The scalability and robustness of this algorithm need to be tested against a wider selection of segregated environments. Real-world constraints such as the suitability of a neighborhood and the quality of a school are also not tackled

by this work. Future research will explore real-world applications of this algorithm, as well as its potential synergies with broader policy initiatives aimed at promoting equitable access to quality education.

Acknowledgements This research was supported by the Innovation Center for AI (ICAI, The Netherlands) and the City of Amsterdam.

References

1. Abdulkadiroğlu, A., Sönmez, T.: Random serial dictatorship and the core from random endowments in house allocation problems. Econometrica **66**(3), 689–701 (1998)
2. Böhlmark, A., Holmlund, H., Lindahl, M.: Parental choice, neighbourhood segregation or cream skimming? an analysis of school segregation after a generalized choice reform. Journal of Population Economics **29**(4), 1155–1190 (2016)
3. Boterman, W., Musterd, S., Pacchi, C., Ranci, C.: School segregation in contemporary cities: Socio-spatial dynamics, institutional context and urban outcomes. Urban Studies **56**(15), 3055–3073 (Nov 2019). https://doi.org/10.1177/0042098019868377, https://doi.org/10.1177/0042098019868377, publisher: SAGE Publications Ltd
4. Boterman, W.R.: Socio-spatial strategies of school selection in a free parental choice context. Transactions of the Institute of British Geographers **46**(4), 882–899 (2021). https://doi.org/10.1111/tran.12454
5. Courtioux, P., Maury, T.P.: Private and public schools: A spatial analysis of social segregation in france. Urban Studies **57**(4), 865–882 (2020)
6. Dignum, E., Athieniti, E., Boterman, W., Flache, A., Lees, M.: Mechanisms for increased school segregation relative to residential segregation: a model-based analysis. Computers, Environment and Urban Systems **93**, 101772 (Apr 2022). https://doi.org/10.1016/j.compenvurbsys.2022.101772
7. Duncan, O.D., Duncan, B.: A Methodological Analysis of Segregation Indexes. American Sociological Review **20**(2), 210–217 (1955). https://doi.org/10.2307/2088328, publisher: [American Sociological Association, Sage Publications, Inc.]
8. Gallagher, M., Lamb, R.: Integrating housing and education solutions to reduce segregation and drive school equity. research report. Urban Institute (2023)
9. Hallinan, M.T.: Diversity effects on student outcomes: Social science evidence. Ohio St. LJ **59**, 733 (1998)
10. Hymel, S., Katz, J.: Designing classrooms for diversity: Fostering social inclusion. Educational Psychologist **54**(4), 331–339 (2019)
11. Johansson, O.: How do independent school admission rules affect school segregation?: An agent-based model in a swedish context (2022)
12. Kemple, J.J.: High school closures in new york city: Impacts on students' academic outcomes, attendance, and mobility. report. Research Alliance for New York City Schools (2015)
13. Michailidis, D., Tasnim, M., Ghebreab, S., Santos, F.P.: Towards reducing school segregation by intervening on transportation networks. Citizen-Centric Multiagent Systems 2023 (CMAS'23) p. 4 (2023)

14. Oosterbeek, H., Sóvágó, S., Klaauw, B.: Why are Schools Segregated? Evidence from the Secondary-School Match in Amsterdam (Jan 2019), https://papers.ssrn.com/abstract=3319783
15. Sage, L., Flache, A.: Can ethnic tolerance curb self-reinforcing school segregation? a theoretical agent based model. arXiv preprint arXiv:2006.13531 (2020)
16. Schelling, T.C.: Micromotives and macrobehavior. WW Norton & Company (2006)
17. Sissing, S., Boterman, W.R.: Maintaining the legitimacy of school choice in the segregated schooling environment of Amsterdam. Comparative Education **59**(1), 118–135 (Jan 2023). https://doi.org/10.1080/03050068.2022.2094580
18. Stoica, V.I., Flache, A.: From Schelling to Schools: A Comparison of a Model of Residential Segregation with a Model of School Segregation. Journal of Artificial Societies and Social Simulation **17**(1), 5 (2014)
19. Wei, R., Feng, X., Rey, S., Knaap, E.: Reducing racial segregation of public school districts. Socio-Economic Planning Sciences **84**, 101415 (2022)

4.2 Agent Interventions to Reduce Procrastination

Agent Interventions to Reduce Procrastination

Ethan Beaird[1], Feyza Merve Hafızoğlu[2], and Sandip Sen[1]

[1] The University of Tulsa, Tulsa OK 74104, USA
[2] İstanbul Commerce University, İstanbul 34840, Turkey

Abstract. Procrastination behavior, given easy access to entertainment options and targeted social media, adversely affects the lives of fellow citizens. This paper presents a model of procrastination on task completion and agent-based interventions to assist citizens in overcoming procrastination. The agent engages the citizen, i.e., user, using instances of given task types to develop a shared awareness of user preferences and capabilities. This preference model is then used to both choose effective interventions as well as measure and reward subsequent user performance.

Keywords: human-agent interaction · procrastination

1 Introduction

Our research is motivated by the goal of using agent assistants to improve the productivity and well-being of our fellow citizens. In particular, agent assistants can aid users in resolving conflicts, eliminating process inefficiencies, and better utilizing native skills, available resources, and existing relationships. In this paper, we describe the use of agent interventions to redress a specific behavioral inefficiency that affects fellow citizens: procrastination.

Procrastination, i.e., irrational delay in completing tasks of importance at hand [14], is caused by failure of self-regulation [13] and lack of self-efficacy [7] and has increasingly posed significant challenges to citizens of our digitally connected world [4]. The disparity between people's intentions to complete important tasks and their tendency to procrastinate illustrates an inclination to irrationally prioritize immediate gratification, i.e. short-term reward over long-term benefits. People tend to procrastinate on necessary tasks and chores that are perceived to be negative, unpleasant, or challenging [16].

Social media companies reward people with instant gratification, and omnipresent communication platforms result in citizens spending considerable time and energy [9] in initiating and responding to messages and calls that can easily divert attention from pending tasks at hand [1].

Considering its negative consequences on individuals' health, well-being [2], and task accomplishments [6], eliminating or reducing procrastination is a challenging but potentially significantly beneficial endeavor. A range of interventions, although not widely available, have been developed in the psychology literature [3] to reduce procrastination levels, such as self-regulation [5], cognitive-behavioral therapy [7], and social support, among others. However, only a minority of those who suffer from procrastination have access to such interventions

due to financial limitations and other resource barriers, and they are not perfect remedies. Leveraging agent technology to address this pervasive issue can prove highly effective, as these automated assistants, serving as personal aids, can be widely deployed at minimal costs, thus helping reduce the inequity and disparity in access to procrastination interventions.

Despite the prevalence and significant impact of procrastination in our society, the majority of previous studies focus only on academic procrastination [13]. Furthermore, most studies rely on surveys rather than actual task performance data. We evaluate two key research questions on the effectiveness of agent interventions to reduce procrastination using actual task performance data:

• Can agent intervention mechanisms be developed to help people significantly reduce procrastination?

• Do different levels of agent interventions have significantly different effects?

We choose four diverse task types that span a range of basic skills and on which we expect different procrastination likelihoods among individuals. To assess the procrastination tendency of individuals for a certain task, we use their preferences among the task types. This is because an individual's likelihood to procrastinate on a task type and their preference for that task type are correlated [10].

We conduct experiments with human subjects to simulate procrastination and design utility functions based both on the participants' preferences for different task types and their performance levels while completing those tasks. We ensured that completing less preferred tasks with lower performance can still yield higher utilities than highly preferred tasks completed with higher performance. Each participant is asked to perform a set of task instances selected from the given task types. This set of task instances is completed with the aid of either *low* or *high level agent interventions*. We collect both task completion metric data and survey data to assess individuals' perceptions of procrastination and agent interventions as well as satisfaction in task completion. Our data analysis corroborates our position that the use of agent-based interventions plays a significant role in helping individuals reduce procrastination.

2 Related Work

Procrastination behavior is an increasingly serious problem in our professional [13] and daily lives [8]. Empirical research on addressing procrastination is scarce, where the majority use survey data [4, 8, 13] and involve college students in academic domains [13]. Recently, there is a growing body of literature that is concerned with adults' procrastination in life-domains [4]. The indisputable effect of online entertainment and other sources has contributed significantly to the rise of procrastination behavior among adults [4].

Numerous factors influencing procrastination are addressed in social literature such as age, gender, personality, mood, environment, and nature of task [16]. It is recognized that procrastination causes lower levels of health, well-being,

and achievements, i.e., performance [6]. Various helpful intervention mechanisms have been identified [3].

Recently, several studies have suggested that technology-based interventions can be useful in overcoming irrational delay. Zavaleta *et al.* [17] showed that email interventions by instructors can reduce the delay of students starting their online homework. De Vries *et al.* [15] found that experts sending motivational messages through a digital medium could be more motivating in the earliest stages of behavior change, while peer-designed messages in a digital medium could be more motivating in the later stages. GanttBot [12] is a chatbot that is developed using conversational agents with several abilities: reminding students about landmarks, informing tutors when interventions are needed, and the ability to learn from previous interactions. StudiCare [11] is a digital coach rooted in internet and mobile-based cognitive behavioral therapy techniques, helping guide students to achieve their academic goals.

The present study extends existing research on procrastination as follows:

Time: Address procrastination in short-term with short artificial tasks rather than long-term [11, 17].

Data: In addition to survey data, use actual task performance metrics.

Task: Task domain involves daily life tasks rather than a specific domain such as in [6, 11–13, 17]

Users: Recruit adult participants beyond only college students [12, 17].

Agents: Make use of cost-effective and ubiquitous agent intervention mechanisms rather than time and location-dependent techniques [7].

3 Procrastination Model

Task Domains: We use four task types—*Audio Transcription, Shopping, Text Comprehension*, and *Form Filling*— to emulate and represent a variety of common online activities.

Audio Transcription task (see Figure 1a) presents users with a brief audio clip and a text box where they are prompted to transcribe the audio. *Shopping* task (see Figure 1b) gives users multiple itemized receipts, each featuring randomly generated prices and a discount rate percentage. Users must select the receipt with the lowest total cost. *Text Comprehension* task (see Figure 1c) presents users a short story and then asks multiple-choice questions to assess their understanding of the text. *Form Filling* task (see Figure 1d) includes a simulated online tax form to be filled by the users based on a simulated ID card.

Utility Function: We adopt the following definition of procrastination: "to voluntarily delay an intended course of action despite expecting to be worse off for the delay" [14]. Milgram et al.[10] observes: "Procrastination is found to be greater in tasks that are regarded as unpleasant." To study procrastination in an online setting, we created conditions to engender

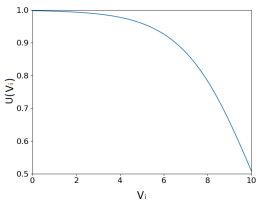
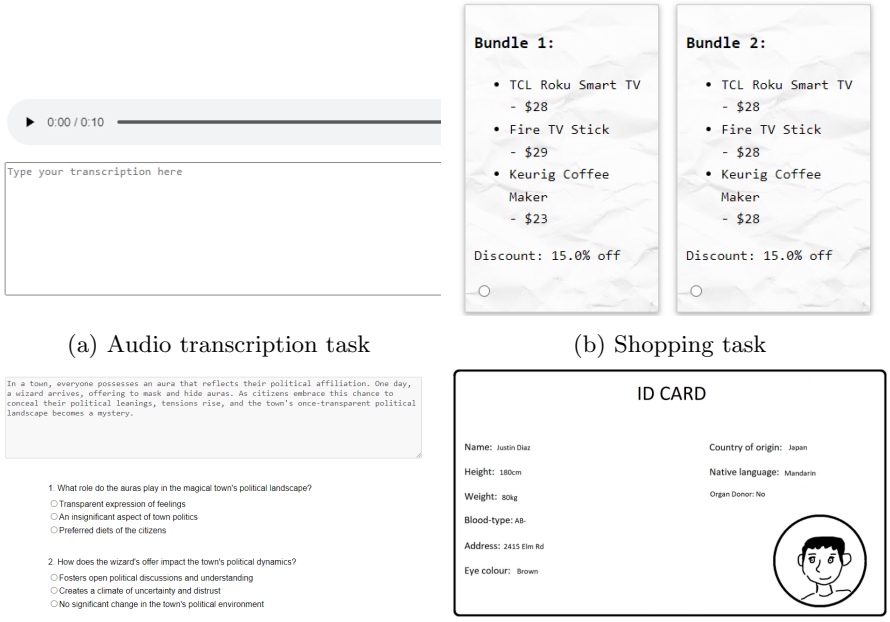


Fig. 2: Utility function



(a) Audio transcription task

(b) Shopping task

(c) Text comprehension task

(d) Form filling task

Fig. 1: Instances of task types used in our study on procrastination.

procrastination. Procrastination in our study corresponds to users selecting preferred tasks of lower utility.

Each user completes initial demo instances of each task type, and we record their scores. Additionally, each user ranks their preference for each task type. The value of task i's completion, v_i, by a user: $v_i = w \cdot 10 \cdot s_i + (1 - w) \cdot p_i, i \in \{1, 2, 3, 4, 5\}$ where $s_i \in [0, 1]$ is the user's score in the initial demo, $p_i \in [1, 10]$ is the preference ranking, and $w = 0.3$ is the weighting of s_i for task type i, chosen to emphasize the user's task preference while still considering the user's ability to perform that task. Then we can construct the utility function over the task types as follows: $U(v_i) = \frac{1}{1+e^{0.625 \cdot v_i - 2\pi}}$.

This utility function is designed to discourage selecting tasks that a user may prioritize by penalizing tasks that align with their preferences and abilities. Utility decreases as preference ranking p or initial task demo score s increases.

Agent Interventions: We present several proactive and reactive agent interventions to redress procrastination.

Motivational Guide is a subtle proactive intervention where the agent gently encourages users to select higher utility tasks through text messages (see Figure 3).

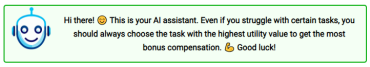


Fig. 3: Motivational guide intervention.

Dynamic Utility Highlighting is a proactive intervention scheme where the agent adds "recommendation" popups to tasks with the highest utility and highlights them with a green outline. Lower utility tasks are outlined in red. Additionally, tasks are sorted in descending order of utility (see Figure 4).

Medal Rewards is a reactive intervention featuring a "medals box" where users can view earned medals after completing a task. Medals are awarded based on task selection and performance. Users earn bronze medals for selecting tasks that are not the highest utility. For selecting high utility tasks, users earn silver and gold medals based on their performance (see Figure 5). This intervention is paired with *Reactive Motivation*, where the agent leaves encouraging comments tailored to the user's performance in the previous task.

Fig. 4: Dynamic utility highlighting intervention.

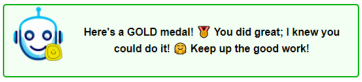


Fig. 5: Medal rewards & reactive motivation intervention.

4 Experimental Methodology

Hypothesis 1: Agent intervention mechanisms can be developed to help people overcome or significantly reduce procrastination.
Hypothesis 2: Low and high level agent interventions can have different effects on reducing procrastination.
Experimental Setup: Three conditions on agent intervention intensity are:
No Intervention group has no agent interventions present.
Low Intervention group includes the *Motivational Guide*, *Medal Rewards*, and *Reactive Motivation* interventions, providing users with involved but minimally intrusive guidance and motivation.
High Intervention group includes the *Motivational Guide* and *Dynamic Utility Highlighting*, offering a more visually prominent form of guidance.

The experiment consists of the following steps:
• *Preference learning:* Users complete a series of demo tasks for each of the four task types and rank their preferences for each task.
• *Task phase:* Users are given twelve task instances— three instances of each of the four task types. They choose any six to complete. Users in the *low* or *high* groups receive agent interventions during this phase. After completing chosen tasks, users take a final survey on satisfaction, agent interaction, and procrastination. Users get one minute to complete each task and are paid proportional to the overall utility from completing tasks.
Metric: We utilize the *Selection score* metric in our analysis. To analyze the impact of agent interventions on a user's task selection and allocation, we define *optimal utility* as $O = 3(U(t_1) + U(t_2))$ where t_1 and t_2 are the first and second highest utility scores. We define *user utility* as $\mu = \sum_T U(t_i)$, where T refers to the tasks selected by the user. Then, *selection score* is $\frac{\mu}{O}$.

We experimented with 180 participants (60 participants for each condition) recruited through Amazon Mechanical Turk.

Results: Figure 6 presents selection scores for all three conditions. In the high intervention condition, the *selection score* ($M = 0.912, SD = 0.108$) is significantly ($p < 0.001$) higher than that in the no intervention condition ($M = 0.840, SD = 0.096$) and the low intervention condition ($M = 0.845, SD = 0.092$). No significant difference in selection scores is found between no and low intervention conditions. These results indicate that high level agent interventions lead to a significant increase in the selection of the less preferred, i.e., procrastinated, tasks. Additionally, responses to the survey item "I liked the outcome of my choices from the main phase's task list" were significantly higher ($p < 0.05$) in the low and high intervention conditions than the no intervention condition. No significant difference in other survey items is found between the conditions.

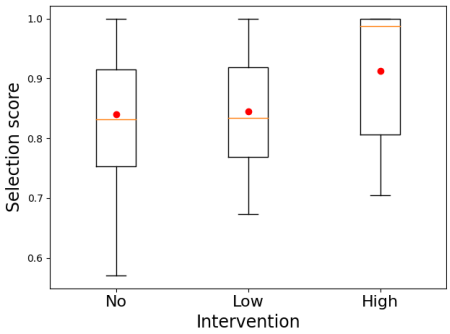


Fig. 6: Comparison of selection scores

5 Discussion & Conclusion

This study investigates the role of agents in controlling procrastination by testing various agent interventions in an online environment. Many users in our experiment preferred to perform tasks that they enjoyed rather than those that return higher utility. These irrational task selections are voluntary, akin to a form of *irrational delay*.

Results show that timely (high level) agent interventions can significantly reduce users' procrastination tendencies as reflected in improved task selection and completion rates. **Hypothesis 1** is partially supported. Furthermore, while only high interventions effectively reduce procrastination tendencies, a statistically significant difference is found between the two conditions that is consistent with **Hypothesis 2**.

In conclusion, our study addresses the pervasiveness of procrastination in a highly connected and overstimulated modern era through the use of novel agent intervention techniques. While many psychological interventions and schools of thought already exist to address procrastination tendencies, these methods are often inequitable and not accessible to all due to financial and other resource limitations. In response, we have developed a model of procrastination through the use of an online environment as well as two agent intervention conditions composed of multiple agent interventions aimed at addressing and mitigating procrastination tendencies in citizens. Our findings suggest that user procrastination can be effectively modeled and that agent interventions can provide useful tools for helping citizens reduce their procrastination tendencies.

References

1. A.A., H., J, K., M, A.: Academic procrastination of medical students: The role of internet addiction. Journal of Advances in Medical Education & Professionalism **8**(2), 83–89 (2020)
2. Constantin, K., English, M.M., Mazmanian, D.: Anxiety, depression, and procrastination among students: Rumination plays a larger mediating role than worry. Journal of Rational-Emotive & Cognitive-Behavior Therapy **36**, 15–27 (2018)
3. van Eerde, W., Klingsieck, K.B.: Overcoming procrastination? a meta-analysis of intervention studies. Educational Research Review **25**, 73–85 (2018)
4. Geng, J., Han, L., Gao, F., Jou, M., Huang, C.: Internet addiction and procrastination among chinese young adults: A moderated mediation model. Computers in Human Behavior **84**, 320–333 (2018)
5. Häfner, A., Oberst, V., Stock, A.: Avoiding procrastination through time management: an experimental intervention study. Educational Studies **40**(3), 352–360 (2014)
6. Kim, K.R., Seo, E.H.: The relationship between procrastination and academic performance: A meta-analysis. Personality and Individual Differences **82**, 26–33 (2015)
7. Krispenz, A., Gort, C., Schültke, L., Dickhäuser, O.: How to reduce test anxiety and academic procrastination through inquiry of cognitive appraisals: A pilot study investigating the role of academic self-efficacy. Frontiers in Psychology **10** (2019)
8. Kroese, F., De Ridder, D., Evers, C., Adriaanse, M.: Bedtime procrastination: introducing a new area of procrastination. Frontiers in psychology (2014)
9. Lozano-Blasco, R., Robres, A.Q., Sánchez, A.S.: Internet addiction in young adults: A meta-analysis and systematic review. Computers in Human Behavior **130**, 107–201 (2022)
10. Milgram, N.A., Sroloff, B., Rosenbaum, M.: The procrastination of everyday life. Journal of Research in Personality **22**(2), 197–212 (1988)
11. Mutter, A., Küchler, A., Idrees, A.R., Kählke, F., Terhorst1, Y., Baumeister1, H.: Studicare procrastination-randomized controlled non-inferiority trial of a persuasive design-optimized internet- and mobile-based intervention with digital coach targeting procrastination in college students. BMC Psychology **11**(273), 1–17 (2023)
12. Pereira, J., Díaz, O.: Struggling to keep tabs on capstone projects: A chatbot to tackle student procrastination. ACM ToCE **22**(1) (2021)
13. Senecal, C., Koestner, R., Vallerand, R.J.: Self-regulation and academic procrastination. The Journal of Social Psychology **135**(5), 607–619 (1995)
14. Steel, P.: The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. Psychological Bulletin **133**(1), 65–94 (2007)
15. de Vries, R.A.J., Zaga, C., Bayer, F., Drossaert, C.H.C., Truong, K.P., Evers, V.: Experts get me started, peers keep me going: Comparing crowd- versus expert-designed motivational text messages for exercise behavior change. In: International Conference on Pervasive Computing Technologies for Healthcare. p. 155–162. PervasiveHealth '17 (2017)
16. Wieland, L.M., Hoppe, J.D., Wolgast, A., Ebner-Priemer, U.W.: Task ambiguity and academic procrastination: An experience sampling approach. Learning and Instruction **81**, 101–595 (2022)
17. Zavaleta Bernuy, A., Han, Z., Shaikh, H., Zheng, Q.Y., Lim, L.A., Rafferty, A., Petersen, A., Williams, J.J.: How can email interventions increase students' completion of online homework? a case study using a/b comparisons. In: 12th International Learning Analytics and Knowledge Conference. p. 107–118. LAK22 (2022)

4.3 Covid-19 in Hospitals Through the Lens of a Citizen-Centric Agent-Based Model

Covid-19 in Hospitals Through the Lens of a Citizen-Centric Agent-Based Model

Philippos Michaelides[1][0009−0002−5505−1588] and
Stefan Sarkadi[2][0000−0003−3999−528X]

[1] School of Business and Economics, Maastricht University, The Netherlands
philippos.michaelides@maastrichtuniversity.nl
[2] Dept. of Informatics, King's College London, London, United Kingdom
stefan.sarkadi@kcl.ac.uk

Abstract. Hospitals are highly dynamic environments where Covid-19 is highly transmissible if effective measures are not taken. Several factors drive its transmission, including mask efficiency (or lack thereof), environment setup, and agent behaviour. General preventive policies are not necessarily effective in specific settings. However, agent-based modelling can offer a way to tailor policies according to various factors that influence the management of institutions, environments, and the activities of citizens within these setups. In this paper, we show how an agent-based model can simulate a typical hospital in a citizen-centric fashion. We customise the study of how Covid-19 is transmitted by focusing on the diverse characteristics of its population. We derive more tailored conclusions regarding influencing factors and, consequently, formulate more precise preventive actions that could enable us to better respond to future outbreaks.

Keywords: Covid-19 · healthcare · multi-agent systems · agent-based modelling · citizen-centric simulation

1 Introduction

Since the start of the Covid-19 pandemic (henceforth 'Covid' or 'Virus' or 'Pandemic'), our lives have gradually returned to their pre-pandemic state. We have seen how the evolving understanding of the factors influencing transmission, and mainly the vaccine distribution, resulted in more effective measures being implemented. Yet, research into Covid continues to this day, with several fundamental medical research papers serving as the basis for subsequent mathematical and agent-based modelling studies. These latest studies aim to simulate the Covid spread within specific environments and situations to evaluate or even propose new customised suggestions for prevention. However, too few of them aim to take a tailored approach to agent-based simulations.

This paper focuses on investigating and assessing the effectiveness of face masks in hospitals. To achieve this, we develop an agent-based model (ABM) that simulates a typical hospital setting, allowing us to quantify the impact of face masks on Covid transmission under different factors.

Previously, agent-based modelling has been employed to study internal environments, such as buildings and facilities, and external environments, such as small cities or entire countries. For example, Baccega et al. employed a model to assess active preventive strategies in a typical school environment during the Pandemic. Their research aimed to identify the most effective control strategy for mitigating the spread and avoiding extensive school closures while adjusting to disease spread [3]. Macalinao et al. conducted another study in which they simulated a typical classroom in the Philippines to investigate the effects of human interactions on Virus transmission in schools [8]. Ying et al. developed a model for evaluating the effectiveness of mitigation strategies in supermarkets. Their model considers customer movement and a transmission model determining if a given time spent close to infected individuals could transmit the Virus. Using this model, the impact of various mitigation strategies on reducing human-to-human transmission in the highly dynamic environment of a supermarket was effectively quantified [12]. Additionally, Ciunkiewicz et al. suggested a configurable model to simulate Covid spread in small, highly localised, and variable environments, namely offices, campuses, or long-term care facilities. The main objective was to provide actionable insights by forecasting transmission while considering epidemiological parameters, airborne viral spread, vaccination, and mask-wearing [4]. Also, researchers have agent-based modelled cities and countries to evaluate the competence of preventive measures implemented by governments and local councils. For instance, Hoertel et al. evaluated the effect of post-lockdown interventions, such as social distancing and usage of masks, based on the cumulative disease incidence, mortality, and ICU-bed occupancy in France [5]. Finally, Wilder et al. applied agent-based modeling to study the impact of demographic structure on Covid transmission in Hubei, Lombardy, and New York City, while also assessing the effectiveness of preventive measures [10,11].

While face masks have already been studied in the context of medical research and mathematical analyses, the benefit of this work is that it applies and adjusts the knowledge gained to the hospital setting. Covid transmission has been studied for isolated single hospital units and simple interactions [6], and our ABM builds upon this research by expanding to realistically encompass a complete, highly dynamic, and diverse hospital environment. It accounts for many of its population and interaction dynamics by simulating and thus considering complex human behaviors and particularities [9]. It is also parameterizable so that it could be distributed to various hospitals around the world, tailored to their specific settings and needs. It could serve as a valuable and efficiently running multi-agent decision-making tool for governments or hospital management systems. Such a tool would benefit citizens around the world by enabling their governing institutions to reach conclusions tailored to specific environments and resource availability in the health and public safety domain. This would lead to more effective public health strategies and enhance the quality of healthcare delivery and social well-being for diverse populations.

The source code for the model and the Appendix can be found on the Open Science Framework (OSF) at the following link: supplementary material.

2 Modelling Covid in Hospital full of Agents

We model a typical medium-sized hospital, emphasising the key facilities, operations, policies and stakeholders. We adapt the Baccega school model to the hospital environment with the aim of addressing the research question: *How and to what extent do face masks influence transmission in an indoor hospital environment?* We model and simulate seven scenarios with different combinations of the factor values using sensitivity analysis. We focus on the type of masks and the percentage of people wearing them, aiming to explore the correlation (positive or negative) with the Virus spread and the magnitude of impact.

As for the layout and structure of the hospital building, apart from the doctor offices and patient rooms with private toilets (double or single, totaling 32 beds), the building includes emergency rooms, surgery rooms, and toilet facilities. Additionally, the hospital features a spacious reception area, a cafe, a small staff kitchen, and the necessary corridors. Regarding the individuals (stakeholders) who utilise the hospital, we can identify three distinct groups (roles): a) patients, who can be either inpatients or outpatients; b) staff, including doctors, nurses, cleaners, and receptionists; and c) visitors. Each group is modeled by considering its diverse characteristics, ensuring that all their dynamic interactions are realistically simulated. The population within the hospital is dynamic, with permanent changes observed in the patient and visitor population due to scheduled medical visits (doctor appointments), emergency medical visits, visits from visitors, admissions and discharges. However, the staff population remains constant, with only temporary changes due to staff shifts. Hospital operating hours are between 07:00 and 19:00 and include admissions, discharges, doctor appointments, emergency medical visits, surgical procedures, visiting hours, and cafe availability, with most hospital staff on duty. Outside these hours, only emergency patient admissions and medical visits are permitted (at lower rates), while the hospital operates with a reduced staff. Finally, while the model primarily focuses on the dynamics within the hospital setting, we do not entirely ignore the external environment. We take into account the possibility that an individual entering the hospital could be infected, based on a fixed probability. The design decisions represent a simplified version of established practices that are followed in real-life cases, and they are fully parameterisable so that the model can adjust and capture various situations [3].

We use the well-known Susceptible-Exposed-Infected-Recovered (SEIR) model [2]. *Susceptible* individuals are those able (vulnerable) to contract the Virus, *exposed* have been infected but are not yet infectious, *infected* have been infected and are infectious, and *recovered* have become immune and are neither infected nor infectious [4].

[3] Additional details of the design and assumptions concerning the environment, population, policies, and human behavior, as well as the external environment, can be found in Appendix A1.

[4] Details of the epidemic model and transmission drivers can be found in Appendix A2.

3 Methodology and Preliminary Results

Table 1 illustrates the seven scenarios simulated [5]. We only account for the N95 medical respirator (henceforth 'N95') and medical grade procedure (henceforth 'surgical') mask types, with efficacy levels of 99% and 59%, respectively [7,1]. We also consider wearing percentages of 0%, 50%, 75% and 100% [6]. We use a baseline scenario in which no mask is worn, which is then independently compared with scenarios where the mask is used under different conditions (in terms of mask type and wearing percentage) to quantify the mask impact under each of these conditions. We conduct hypothesis tests, aiming to determine whether there is a statistically significant difference between the statistics (means) resulting from the different scenarios. We make our decision based on the average daily generated cases of each scenario.

Table 1: Experimental set-up. Scenario 1 assumes no mask. Each Scenario 2-7 simulates a different mask-wearing condition regarding mask type and wearing percentage.

	Mask type		***Mask-wearing percentage(%)***				
	N95	**Surgical**	**No**	**0**	**50**	**75**	**100**
1			x	x			
2	x				x		
3	x					x	
4	x						x
5		x			x		
6		x				x	
7		x					x

We selectively present and visually highlight the important metrics and statistics based on the means (and the standard deviations) of the total of 30 independent simulation runs performed for each scenario.

To investigate mask impact regarding type and wearing percentage, all mask-wearing conditions were compared to the baseline scenario of no mask. All the t-tests returned very low p-values (lower than our alpha of 0.05), meaning there is a statistically significant difference in Covid spread between wearing and not wearing a mask for all mask types and wearing percentages.

Based on the results summarised in Table 2, Covid spread is significantly reduced in all scenarios where the mask is worn, compared to the no-mask scenario, with this to be reflected in all related metrics and statistics. Namely, when the hospital population wears masks, regardless of the mask type and wearing percentage, the hospital experiences significantly fewer average daily and cumu-

[5] Model implementation and verification details can be found in Appendix A3.
[6] Details of the mask-wearing assumptions can be found in Appendix A4.

lative generated cases and, therefore, a smaller transmission ratio and cumulative cases percentage.

Table 2: Metrics and statistics for Scenarios 1-7. The numbers in parentheses express the percentage changes in average daily generated cases when switching from Scenario 1 to each of Scenarios 2-7.

	No		**N95**			**Surgical**	
	0%	50%	75%	100%	50%	75%	100%
	Scn.1	**Scn.2**	**Scn.3**	**Scn.4**	**Scn.5**	**Scn.6**	**Scn.7**
Average daily generated cases	24.7	11.9 (-52%)	5.1 (-79%)	0.1 (-100%)	16.6 (-33%)	13.0 (-47%)	9.2 (-63%)
Cumulative generated cases	819	392	168	3	553	427	302
Cumulative cases[a]	914	493	273	116	648	528	400
Cumulative cases percentage	10.9%	5.9%	3.2%	1.4%	7.7%	6.3%	4.8%
Transmission ratio	8.6	3.8	1.6	0.0	5.8	4.3	3.1

[a] Includes both generated and exogenous cases (infectious entrants).

Specifically, not wearing a mask could infect more than 10% of the population, reporting around 25 daily generated cases. This is also reflected by the transmission ratio of about nine, meaning that for every single infectious entrant, nine vulnerable individuals are eventually infected. Mask-wearing is an effective measure, resulting in fewer generated cases and a lower transmission ratio, mitigating the spread inside the hospital even as infectious individuals continue to arrive at a constant rate. When worn at a low percentage (50%), the surgical mask shows a significant decrease in spread of around 30%. Higher percentages, especially with N95 masks, can reduce the spread by more than 50%. We highlight the impressive total absence of generated cases when 100% of the population wears N95 masks; the 113 infectious entrants lead to only three infections in 31 days (transmission ratio of just 0.03). All percentage changes are shown in Table 2. Figure 1 represents the comparative results between the baseline scenario and Scenarios 4 and 7, where the N95 and surgical masks, respectively, are worn 100%, showing the significant impact on the cumulative generated and daily generated cases.

4 Conclusions

In this paper we showed how a citizen-centric agent-based modelling approach can be applied to study the spread of Covid within a typical hospital. The approach allows us to reach tailored conclusions, driven by an enhanced understanding of the Virus spread in complex and dynamic environments. We examined the effectiveness of face masks and concluded that they have a significant impact. We found that mask-wearing is a crucial preventive measure, given the

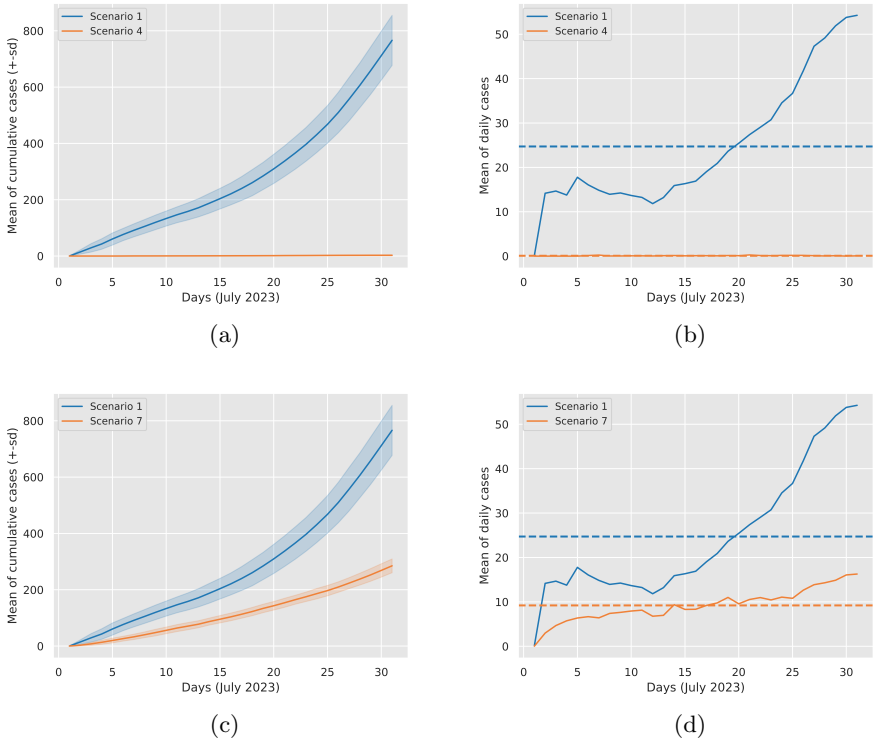


Fig. 1: Comparison of the impact between N95 (Scenario 4) and surgical masks (Scenario 7) at 100% mask-wearing percentage. The figure displays the cumulative generated cases per day and the daily generated cases. In Sub-figures (a) and (c), the solid line represents the 30-run mean of the cumulative generated cases per day, and the colourful area around it is the standard deviation. In Sub-figures (b) and (d), the solid line represents the 30-run mean of the daily generated cases, and the dashed line is the average.

significant reduction in reported cases, especially when the N95 mask is worn continuously by the entire population. Wearing surgical masks of lower efficacy significantly also reduces transmission, but to a relatively limited extent.

The model is somehow limited by the empirical findings of other models that drive the simulations of our transmission mechanism. However, adopting such models has enabled us to make our simulations more realistic and obtain more accurate results. As future work, we would be able to explore additional influencing factors, such as vaccination rate, social distancing, and room ventilation, so that preventive policies could be even better tailored to citizens. Moreover, further parameterising the model would allow its tailoring to other infectious diseases.

References

1. Types of Masks and Respirators. Centers for Disease Control and Prevention https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/types-of-masks.html (2023), accessed: April 12, 2024
2. Aron, J.L., Schwartz, I.B.: Seasonality and period-doubling bifurcations in an epidemic model. Journal of theoretical biology **110**(4), 665–679 (1984)
3. Baccega, D., Pernice, S., Terna, P., Castagno, P., Moirano, G., Richiardi, L., Sereno, M., Rabellino, S., Maule, M., Beccuti, M., et al.: An Agent-Based Model to Support Infection Control Strategies at School. JASSS **25**(3), 1–15 (2022)
4. Ciunkiewicz, P., Brooke, W., Rogers, M., Yanushkevich, S.: Agent-based epidemiological modeling of COVID-19 in localized environments. Computers in Biology and Medicine **144**, 105396 (2022)
5. Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Rico, M.S., Limosin, F., Leleu, H.: A stochastic agent-based model of the SARS-CoV-2 epidemic in France. Nature medicine **26**(9), 1417–1421 (2020)
6. Huang, Q., Mondal, A., Jiang, X., Horn, M.A., Fan, F., Fu, P., Wang, X., Zhao, H., Ndeffo-Mbah, M., Gurarie, D.: Sars-cov-2 transmission and control in a hospital setting: an individual-based modelling study. Royal Society open science **8**(3), 201895 (2021)
7. Lindsley, W.G., Blachere, F.M., Law, B.F., Beezhold, D.H., Noti, J.D.: Efficacy of face masks, neck gaiters and face shields for reducing the expulsion of simulated cough-generated aerosols. Aerosol Science and Technology **55**(4), 449–457 (2021)
8. Macalinao, R.O., Malaguit, J.C., Lutero, D.S.: Agent-Based Modeling of COVID-19 Transmission in Philippine Classrooms. Frontiers in Applied Mathematics and Statistics **8** (2022)
9. Stein, S., Yazdanpanah, V.: Citizen-centric multiagent systems. In: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems. pp. 1802–1807 (2023)
10. Wilder, B., Charpignon, M., Killian, J.A., Ou, H.C., Mate, A., Jabbari, S., Perrault, A., Desai, A., Tambe, M., Majumder, M.S., et al.: The role of age distribution and family structure on covid-19 dynamics: A preliminary modeling assessment for hubei and lombardy. Available at SSRN **3564800** (2020)
11. Wilder, B., Charpignon, M., Killian, J.A., Ou, H.C., Mate, A., Jabbari, S., Perrault, A., Desai, A.N., Tambe, M., Majumder, M.S.: Modeling between-population variation in covid-19 dynamics in hubei, lombardy, and new york city. Proceedings of the National Academy of Sciences **117**(41), 25904–25910 (2020)
12. Ying, F., O'Clery, N.: Modelling COVID-19 transmission in supermarkets using an agent-based model. Plos one **16**(4), e0249821 (2021)

4.4 Effect of Task Allocation Protocols in Human-Agent Teams

Effect of Task Allocation Protocols in Human-Agent Teams

Sami Abuhaimed, Selim Karaoglu, and Sandip Sen

The University of Tulsa, Tulsa OK 74132, USA
{saa8061,sek6301,sandip}@utulsa.edu

Abstract. Ad hoc human-agent teams will be more common in dynamic environments where team members interact without prior experience and only for a limited number of interactions. Human teammates' satisfaction with the task allocation mechanism is also critical for team viability. The team's task allocation mechanism must harness the diverse capabilities of its members for optimal team performance. We study task allocation protocols with varying degrees of flexibility and human control: (a) alternating, (b) performance adaptive, (c) agent-guided, and (d) human-selected. We evaluate the relative strengths of these task allocation procedures for team performance and human satisfaction through experiments with MTurk workers.

Keywords: Task Allocation · Team Performance · User Satisfaction.

1 Introduction

Agents can collaborate with people on critical tasks, such as guiding emergency evacuations [16] and disaster relief [15]. Intelligent agent applications typically assume human roles in human-agent teams, e.g., tutor [18] and trainer [12]. Researchers are studying the interactions and dynamics within these teams to improve their design [5].

We are interested in human-agent collaboration in *ad hoc teams* where team members do not have prior knowledge or interaction experience with their teammates [4]. We consider ad hoc teams that try to accomplish tasks chosen from diverse task types. We assume that different human teammates will have different competencies and expertise in various task types. To optimize the performance of a given human-agent team, allocating tasks to the teammates based on their relative expertise is necessary.

The allocation problem is challenging because a team member does not have *a priori* knowledge of the levels of expertise of its partner. Although we allow human and agent partners to share their estimated expertise over different task types, the accuracy and consistency of such estimates might be unreliable [7]. Repeated interaction allows partners to refine the initial estimates provided. Still, such opportunities are few because (i) only a limited number of repeated teamwork episodes and (ii) allocation decisions that determine what task types are performed by a partner in an episode. Therefore, the success of such ad

hoc human-agent teams in completing tasks will critically depend on effective adaptability in the task allocation process.

Previous work has shown that agent allocators produce higher team performance than human counterparts [1]. However, human satisfaction with the protocol was low [2]. Human teammates may like more control/input of the task allocation process. Thus, the main question this paper studies is: How is the effectiveness of ad hoc human-agent teams influenced by varying levels of responsibilities and input with respect to task allocation? We design four task allocation protocols with various degrees of responsibilities and input from the agent and human team members: Performance-based, Turn-taking, Agent-guided, and Human-selected. We present results from experiments with MTurk workers that involve two key metrics, team performance and human satisfaction in teammates, that measure the viability of such human-agent ad hoc teams.

2 Related Work

Human-agent teams are studied in many domains, such as space robotics [5]. Most of this work is on agents in supportive roles to humans [9] and in robotic and simulation settings [17]. We focus on ad hoc environments, where other work assumes interactions with agents and environments before the study [5]. We are also interested in autonomous agents as "team members fulfilling a distinct role in the team and making a unique contribution" [10].

Task allocation has been extensively studied in multi-agent teams [8]. Agent teams focus on designing efficient mechanisms for agents to distribute tasks within their society. Task allocation is also studied in the literature on human teams and organizations [14]. However, we are not aware of prior examination of autonomous agents with task allocation roles, compared to humans, in virtual and ad hoc human-agent teams.

Organizations must solve four universal problems, including task allocation, to achieve their goals [14]. The task allocation mechanism, including capability identification, role specification, and task planning, is considered an important component of teamwork [13]. In human teams, the focus is on understanding the characteristics of teams to design the best possible task allocation mechanism. Research on the effect of autonomous agent task allocators on human teams is scarce.

3 Allocation Protocols

We designed four new protocols involving both teammates' allocation capabilities over several episodes.

Fig. 1. Turn-taking Protocol.

Turn-taking Allocator Protocol is where the human and the agent take turns for the allocation role (Fig 1). Before the episode starts, team members share their confidence levels for each task type. The human teammate is tasked with allocating

tasks in the first episode. For the following episodes, the allocator role is alternated between the human and the agent. Alternating the allocator role aids the team members in learning about their partner's allocation capabilities.

Performance-based Allocator Protocol also allows both team members to allocate tasks (Figure 2). The human and the agent allocate tasks in episodes 1 and 2, respectively (A_h^1 and A_a^2). For the remaining episodes, the protocol assigns the allocator role based on average team performance when human ($\overline{TP}_h$) and agent ($\overline{TP}_a$) team members are allocating. For any episode after the first two episodes, the team members whose allocation(s) resulted in higher average team performance are selected as the allocator.

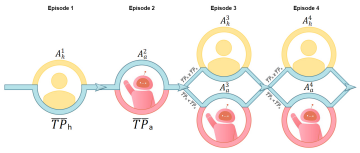


Fig. 2. Performance-based.

Agent-guided Allocator Protocol gives the allocator role to the human for every episode (Figure 3). The agent's role in this protocol is to provide suggestions in the n^{th} episode to the human before allocation. The agent provides task allocation suggestion, $\hat{A}_{ah}^n$, to the human by comparing the human teammate's expressed confidence levels and prior episode team performances, based on which the human determines an allocation, A_h^n .

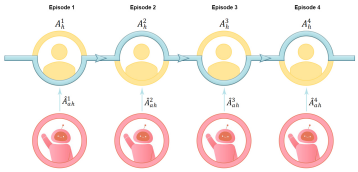


Fig. 3. Agent-guided.

Human-selected Allocator Protocol is when the human team member select the allocator in each episode (Figure 4). Before the start of episode n, the protocol asks the human to select the allocator for nth episode. If the human member chooses to allocate, they allocate task tasks for the episode (A_h^n); otherwise, the agent allocates the tasks (A_a^n).

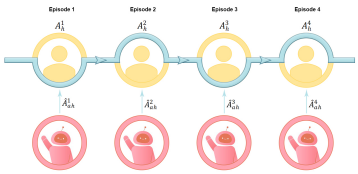


Fig. 4. Human-selected.

4 Hypotheses

H1*: The Performance-based protocol will have the highest team performance.*
H2*: Human satisfaction with the agent will be highest with Agent-guided protocol.*
H3a*: Agent-guided protocol will produce higher team performance than Turn-taking protocol.*
H3b*: Human allocators will increasingly follow agent guidance in the Agent-guided protocol in later episodes.*
H4a*: Humans will select the agent as an allocator more often than themselves.*
H4b*: The Human-selected protocol will produce higher team performance than Agent-guided protocol.*

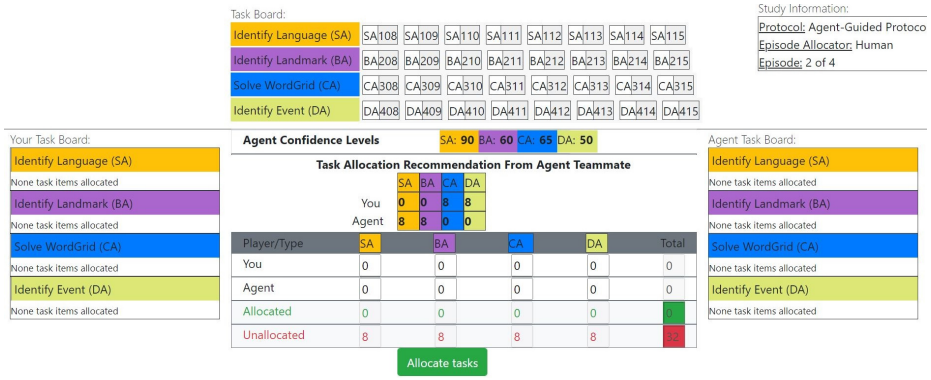


Fig. 5. CHATboard showing allocated tasks to teammates.

H5*: Human-selected and Agent-guided Protocols will have a higher satisfaction than Turn-taking and Performance-based Protocols.*

5 Methodology

Agent Expertise: An agent has a fixed profile that specifies its expertise levels for different tasks, represented as a vector of probabilities for successful completion of task types.
Agent Allocator Strategy: The primary allocation goal is to maximize the use of team capacity, given the expertise of team members. The agent uses estimates of human teammates' task completion rates by task types in the allocation procedure that solves this constrained optimization problem:

$$Max \sum_{y \in M} (x_y a(y) + (1 - x_y)h(y)); s.t.\ \forall y, x_y \in \{0, 1\} \quad (1)$$

where x_y is a binary variable indicating whether a task type, y, is assigned to a human or agent based on the current performance estimate of the human, $h(y)$, and agent, $a(y)$, on task type y.

5.1 Experimental configurations

We use CHATboard, an environment facilitating human-agent, as well as human-human, team collaboration (see Figure 5). We conducted experiments with teams of one human and one agent ($n = 2$), $N = \{p_a, p_h\}$. We use four task types ($m = 4$), M: $\{y_1,y_2,y_3,y_4\}$, which are *Identify Language*, *Solve WordGrid*, *Identify Landmark*, and *Identify Event* (examples of task types shown in Figure 6). The task types are selected so that for each type, sufficient expertise variations in recruited human subjects are likely. We created 32 ($r = 8$) task instances for each of the four episodes ($E = 4$). We recruited 260 participants from Amazon Turk, 65 for each condition (each protocol), as recommended for a medium-sized

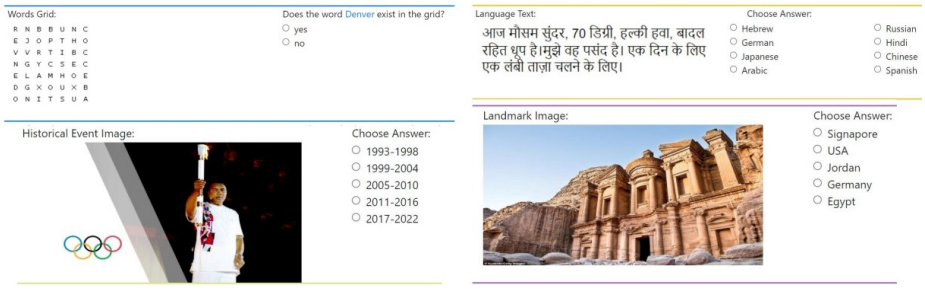


Fig. 6. Task types.

effect [3]. We use a between-subject experimental design, and each team is randomly assigned to a protocol. We incorporate random comprehension attention checks to ensure the fidelity of the result [6].

5.2 Evaluation Metrics

Team Performance: Team Performance is measured as the average team performance over episodes, $\frac{1}{E}\sum_{e=1}^{E} R_{team,e}$, where $R_{team,e}$ is the team performance in episode e, which is the average performance, μ, of all team members, $|N|$, overall task types, $|M|$, in that episode:

$$R_{team,e} \leftarrow \frac{1}{|M||N|}\sum_{i=1}^{|N|}\sum_{j=1}^{|M|}\mu_{i,y_j,e}. \tag{2}$$

Human Satisfaction: We measure the satisfaction with the agent through a satisfaction survey adapted from [11] with five questions. The survey follows a 5-point Likert scale setting administered at the end of the study. We present a sample question from the survey: *"I am satisfied with my agent teammate."*

6 Experimental Results

Team Performance: *Performance-based protocol has the highest performance* ($M_p = 0.76, SD_p = 0.07$) of four protocols. However, an ANOVA shows the performance advantage is not statistically significant ($F = 1.2, p > 0.05$). The performance of Agent-guided protocol ($M_p = 0.73, SD_p = 0.06$) is slightly lower than Turn-taking protocol ($M_p = 0.74, SD_p = 0.07$), and a t-test shows this difference is not statistically significant ($t = 0.5, p > 0.05$). The performance of the Human-selected protocol ($M_p = 0.74, SD_p = 0.05$) is a little higher than the Agent-guided protocol ($M_p = 0.73, SD_p = 0.06$), and a t-test shows difference is not statistically different ($t = 0.6, p > 0.05$).
Satisfaction: Human teammates have the highest satisfaction with the agent when using the Agent-guided protocol ($M_p = 3.86, SD_p = 0.8$); ANOVA shows that the difference is not statistically significant, $F = 1.3, p > 0.05$.

Agent Selection and Guidance: We found that human teammates select the agent as team task allocator more frequently, 53.9% to 46.1%, compared to selecting themselves. We also found that human teammates follow the allocation suggestion from the agent teammate differently for each task type.

Table 1. Protocols Satisfaction and Performance.

Protocols \ Metrics	Satisfaction	Performance
Turn-Taking	3.79	0.74
Performance-based	3.63	**0.76**
Agent-Guided	**3.86**	0.73
Human-Selected	3.61	0.74

7 Discussion and Future Work

Previous work showed that agent allocators outperform their human counterparts [1]. The Performance-based protocol assigns the role based on allocator performance, explaining why it produces the highest performance (**H1**). In addition, since in the Agent-guided protocol, the agent suggests optimal allocations, and in the Turn-taking protocol, the allocator role is alternated, we expected the human allocator to follow agent guidance more and thus produce higher performance than the fixed role alternations. However, we do not observe this difference in performance (**H3a**). This is likely because the human prefers to explore and does not follow the agent's guidance closely. We observe the human allocator over/under-allocate different task types to themselves relative to the agent's suggestions (**H3b**).

In Human-selected protocol, the human selects the agent to allocate tasks but cannot interfere with the agent's optimal allocation decisions. In contrast, the human allocators may choose not to follow the agent's well-informed guidance in the Agent-guided protocol. Therefore, we expected the Human-selected protocol to perform better (**H4b**). We observe a small difference in performance caused by the human selecting the agent teammate as the allocator more often and increasingly more frequently in later episodes (**H4a**). We expect this difference in performance to increase if the teammates interact for more episodes. As the agent can have higher input and control in both the Human-selected and Agent-guided protocols, we expected higher performance with those protocols than with the Turn-taking and Performance-based protocols (**H5**). However, we observed that the Performance-based protocol produced higher performance. This is likely because the first two protocols give human teammate option to explore, which they choose to do, whereas that option is unavailable in Performance-based protocol.

Results also show that Agent-guided protocol produces the highest satisfaction ratings since it gives full control to the human while allowing the agent to contribute by providing suggestions. This implies that it is possible to increase human satisfaction without compromising team performance (**H2**).

An interesting future work would be to evaluate different configurations of agent-guided allocations, i.e., whether humans follow agents providing incorrect suggestions and how dynamics change if allocation suggestions are coming from another human vs an agent.

References

1. Abuhaimed, S., Sen, S.: Effective task allocation in ad hoc human-agent teams. In: HHAI2022: Augmenting Human Intellect, pp. 171–183. IOS Press (2022)
2. Abuhaimed, S., Sen, S.: Human satisfaction in ad hoc human-agent teams. In: International Conference on Human-Computer Interaction. pp. 207–219. Springer (2023)
3. Brinkman, W.P.: Design of a questionnaire instrument. In: Handbook of mobile technology research methods, pp. 31–57. Nova Publishers (2009)
4. Genter, K., Agmon, N., Stone, P.: Role-based ad hoc teamwork. In: Proceedings of the Plan, Activity, and Intent Recognition Workshop at the Twenty-Fifth Conference on Artificial Intelligence (PAIR-11) (August 2011)
5. Gervits, F., Thurston, D., Thielstrom, R., Fong, T., Pham, Q., Scheutz, M.: Toward genuine robot teammates: Improving human-robot team performance using robot shared mental models. In: AAMAS. pp. 429–437 (2020)
6. Hauser, D., Paolacci, G., Chandler, J.: Common concerns with mturk as a participant pool: Evidence and solutions. (2019)
7. Kahneman, D.: Thinking, fast and slow. Macmillan (2011)
8. Korsah, G.A., Stentz, A., Dias, M.B.: A comprehensive taxonomy for multi-robot task allocation. The Intl Journal of Robotics Research **32**(12), 1495–1512 (2013)
9. Lai, V., Tan, C.: On human predictions with explanations and predictions of machine learning models: A case study on deception detection. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. pp. 29–38 (2019)
10. Larson, L., DeChurch, L.A.: Leading teams in the digital age: Four perspectives on technology and what they mean for leading teams. The Leadership Quarterly **31**(1), 101377 (2020)
11. Lee, S., Choi, J.: Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies **103**, 95–105 (2017)
12. Lin, R., Gal, Y., Kraus, S., Mazliah, Y.: Training with automated agents improves peoples behavior in negotiation and coordination tasks. Decision Support Systems (DSS) **60**(1–9) (April 2014)
13. Mathieu, J.E., Rapp, T.L.: Laying the foundation for successful team performance trajectories: The roles of team charters and performance strategies. Journal of Applied Psychology **94**(1), 90 (2009)
14. Puranam, P., Alexy, O., Reitzig, M.: What's "new" about new forms of organizing? Academy of Management Review **39**(2), 162–180 (2014)
15. Ramchurn, S.D., Huynh, T.D., Ikuno, Y., Flann, J., Wu, F., Moreau, L., Jennings, N.R., Fischer, J.E., Jiang, W., Rodden, T., Simpson, E., Reece, S., Roberts, S.J.: Hac-er: A disaster response system based on human-agent collectives. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems. pp. 533–541. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2015)
16. Robinette, P., Wagner, A.R., Howard, A.M.: Building and maintaining trust between humans and guidance robots in an emergency. In: AAAI Spring Symposium: Trust and Autonomous Systems. pp. 78–83. Stanford, CA (March 2013)
17. Rosenfeld, A., Agmon, N., Maksimov, O., Kraus, S.: Intelligent agent supporting human–multi-robot team collaboration. Artificial Intelligence **252**, 211–231 (2017)
18. Sanchez, R.P., Bartel, C.M., Brown, E., DeRosier, M.: The acceptability and efficacy of an intelligent social tutoring system. Computers & Education **78**, 321–332 (2014)