Social Identities and Responsible Agency

Extended Abstract

Karthik Sama
International Institute of Information Technology, Bangalore
Bangalore, India
sai.karthik@iiitb.ac.in

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

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

ABSTRACT

Social identities play an important role in the dynamics of human societies, and it can be argued that some sense of identification with a larger cause or idea plays a critical role in making humans act responsibly. Often social activists strive to get populations to *identify* with some cause or notion– like green energy, diversity, etc. in order to bring about desired social changes. We explore the problem of designing computational models for social identities in the context of autonomous AI agents. For this, we propose an agent model that enables agents to *identify* with certain notions and show how this affects collective outcomes. We also contrast between associations of identity with rational preferences. The proposed model is simulated in an application context of urban mobility, where we show how changes in social identity affect mobility patterns and collective outcomes.

KEYWORDS

Agency; Identity; Multi-Agent Systems

ACM Reference Format:
Karthik Sama, Jayati Deshmukh, and Srinath Srinivasa. 2024. Social Identities and Responsible Agency: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

1 INTRODUCTION

Most systemic changes are feasible when a large number of people participate and contribute in bringing the change. For example, in Amsterdam, cycling accounts for 38% of all vehicle trips, and there are about 0.75 bikes per inhabitant [1]. In order to motivate people, social identity plays a crucial role. When people identify with a cause or notion, they willingly and actively participate in bringing a change. Hence, social identity[6] is a way to encourage responsible behaviour in humans.

In this paper, we propose a model for autonomous agents[5] which have a social identity such that they can *identify* with abstract notions to mimic the social identity of humans. We use one of the existing models of responsible identity, based on the idea of an elastic sense of self, called Computational Transcendence (CT) [2] and extend it so that autonomous agents can identify with multiple abstract notions and act such that their actions are aligned with the notions they identify with.

We demonstrate this model in a scenario where autonomous agents must make transit choices between private and public transport [4]. Autonomous agents identify with different notions like environmentalism, frugalism, etc., each of which impacts their choices. We then study the collective behaviour that emerges in such populations and present the results.

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

2 MODELLING IDENTITY WITH ABSTRACT NOTIONS

In this section, we elaborate on how we model autonomous agents that identify with notions by extending the idea of Computational Transcendence (CT) [2]. The CT framework defines an autonomous agent a with an elastic sense of self. Formally, this elastic sense of self is represented by $S(a) = (I_a, d_a, \gamma_a)$ where:

- I_a represents the set of objects or external entities with which the agent a identifies itself.
- $d_a : a \times I_a \mapsto \Re^+$ is a set of semantic distances.
- $\gamma_a \in [0, 1]$ represents the elasticity or transcendence level of the agent a's sense of self.

For extending CT, we need to differentiate between two entities - *observables* and *objects of identity set*. The measurable quantities in an agent's given environment or context are observables. If the context changes, the observables change correspondingly. However, the objects in the identity set of the agent, as the name suggests, are a part of the agent's identity and thus are independent of the environment or context in which the agent operates. Since abstract notions are part of the agent's identity set, they are also invariant of the context in which the agent operates. Thus, to build autonomous agents that identify with abstract notions, we introduce *schema* to translate observables into identity associations.

The schema of the identity object is defined as the normalised weights of the relevant objects in the identity set over all the observables in a given context $\bar{co} = (co_1, co_2, ...co_n)$. For a relevant identity object o_i in I_a of the agent a, the schema of o_i is defined as follows:

$$\bar{s_{o_i}} = (s_1^i, s_2^i, ...s_n^i) \text{ where } \Sigma_{j=1}^n s_j^i = 1 \tag{1}$$

$$\bar{p} = \frac{1}{\Sigma_{i=1}^m \gamma^{d(o_i)}} \begin{pmatrix} \gamma^{d(o_1)} & \gamma^{d(o_2)} & \dots & \gamma^{d(o_m)} \end{pmatrix} \begin{pmatrix} \bar{s_{o_1}} \\ \bar{s_{o_2}} \\ . \\ \bar{s_{o_m}} \end{pmatrix} \tag{2}$$

Let the identity set of a be $I_a : \{o_1, o_2, o_3..., o_m\}$ having m relevant objects in the given context. Subsequently let each identity object o_i have its corresponding schema $\bar{s_{o_i}} = (s_1^i, s_2^i, ...s_n^i)$. Then, the

preference vector $\bar{p}$ over the observables in that context is defined using the CT framework as follows:

Given a performable action or choice a_i that results in a set of observables in the given context as $\bar{co}_{a_i} : (co_1^i, co_2^i, ..., co_n^i)$, the utility of a_i can be computed as follows, $\bar{p}$ being $(p_1, ..., p_n)$:

$$u(a_i) = \bar{p}.u(\bar{co}_{a_i}) \tag{3}$$

The behaviour of an agent is modelled using the standard Markov Decision Process framework with the sense of self of a transcended agent forming the basis of its behaviour. The semantic distance updates that signify the adaptive capability of the agents depending on the changes in the environment can be calculated as per the original CT framework, since we have specified the process to translate the utility of observables from the environment into the utility of objects in the identity set of an autonomous agent using schema.

3 MODELLING TRANSIT CHOICES

Having defined the formal model of autonomous agents with an elastic sense of self which can identify with notions, next, we demonstrate modelling transit choices as a realistic use-case of this model. Various factors influence the transit choices taken by humans of which we consider cost, time, congestion, and carbon footprint as the relevant contextual observables. Social factors like conformity are also introduced in the model.

Prospect theory [3] is used to model the utility function for the observables– time, cost, and carbon footprint. For the observable congestion, the utility function is modelled as a discontinuous ReLU function.

Currently, we have modelled two transit modes, taxi (private transport) and bus (public transport). However, other modes of transport can also be modelled using this framework. The cost of transit for each mode is a pre-determined constant. A right-skewed distribution, namely the Gumbel distribution is used to model the time taken by the vehicles. The occupancy of a vehicle is used to calculate the observables - congestion and carbon footprint per head. A Gaussian distribution is used to model the occupancy of a bus. In the case of taxis, a discrete probability distribution is used to model the occupancy. The total emission is calculated based on the estimates of the carbon emissions of vehicles.

The relevant notions that the agents identify with, in the context of making transit choices are Frugalism, Idealism, Individualism and Pragmatism. A heuristic schema mapping these objects in the identity set with the observables discussed above is presented in Table 1

We also model conformity among the agents based on the network structure. Let $frac\_neigh_{c_i}$ be the fraction of neighbours who take the choice c_i. Then, an additional utility component is added in the utility computation of choice c_i as follows:

$$u(c_i) += cf * frac\_neigh_{c_i} \tag{4}$$

Notion \ Observable	Cost	Time	Congestion	Carbon FP
Frugalism	$^3/_5$	$^1/_5$	$^1/_5$	0
Idealism	$^1/_{10}$	$^1/_{10}$	$^1/_{10}$	$^7/_{10}$
Individualism	$^2/_{10}$	$^3/_{10}$	$^5/_{10}$	0
Pragmatism	$^1/_3$	$^1/_3$	$^1/_3$	0

Table 1: Schemas of notions over observables

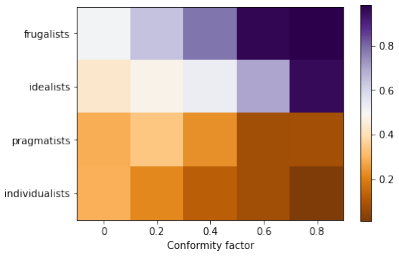

Figure 1: Heatmap of the effect of conformity and semantic distance initialization on population's transit choice

4 EXPERIMENT RESULTS

We generated an Erdős–Rényi network with 500 nodes representing agents, with an average degree of 10. We varied different initial parameters, like the distribution of the initial semantic distances, the amount of conformity in the network, etc. to study the impact of these parameters on the transit choices of the agents. A network of agents is *stabilized* if semantic distances are updated for less than 1% of the agents in the network.

The heatmap in Figure 1 shows the relation between the initial semantic distance distribution and the extent of conformity among the agents in the network. We observe that with an increase in the extent of conformity in the network, the polarity or the strength of transit choices of the population increases. Further, notions like Frugalism and Individualism emerge as strong indicators for behavioural choices made by the agents, while Idealism and Pragmatism are weak indicators.

5 CONCLUSIONS

This extended CT model can help us identify the kind of identity associations that lead to desirable emergent behaviours at a population level. Policymakers and system designers can further use it to understand and design interventions in the system in order to achieve specific goals, for instance in the context of urban mobility– reducing carbon footprint, improving the efficiency of transit, etc.

From the theoretical front, our work helps bridge the gap between abstract notions and real-world observables using our proposition of schemas. While it is difficult to estimate the utility of abstract notions directly, breaking them down as schemas over the observables enables us to compute and estimate their utilities. In turn, this helps build autonomous agents with an identity that can dynamically adapt.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Ralph Buehler and John Pucher. 2009. Cycling to sustainability in Amsterdam. (2009).

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

[3] Daniel Kahneman and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47, 2 (1979), 363–391.

[4] Karthik Sama, Jayati Deshmukh, and Srinath Srinivasa. 2024. Transcending To Notions. *arXiv Preprint* (2024). https://doi.org/10.48550/arXiv.2401.12159

[5] Srinath Srinivasa and Jayati Deshmukh. 2022. AI and the Sense of Self. *arXiv preprint arXiv:2201.05576* (2022).

[6] Henri Tajfel and John C Turner. 2004. The social identity theory of intergroup behavior. In *Political psychology*. Psychology Press, 276–293.