TRANSCENDING TO NOTIONS

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

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

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

Arpitha Malavalli
International Institute of Information Technology, Bangalore
arpitha.malavalli@iiitb.ac.in

Srinath Srinivasa
International Institute of Information Technology, Bangalore
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

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 [4]. 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 is a way to encourage responsible behaviour in humans.

Humans identify with abstract notions and act so as to uphold these notions. For example, people who identify with environmentalism call themselves environmentalists and act appropriately across different contexts, like preferring environment-friendly transit options, buying eco-friendly products, reducing and recycling waste, etc. Thus, identification with a notion affects the choices and actions of the person across multiple scenarios.

Most models of autonomous agents replicate and build on a few specific desired characteristics of humans into the agents. For example, chatbots are modelled to respond like humans, agents in the healthcare domain are designed to be empathetic, etc. We focus on designing agents which can be intrinsically responsible. Specifically, just like humans have a social identity which makes them identify with abstract notions and act responsibly, we propose to design autonomous agents with a sense of self such that they can identify with abstract notions relevant to the system in which they are operating.

In this paper, we propose a model for autonomous agents which have a social identity such that they can *identify* with abstract notions. We use one of the existing models of responsible identity called Computational Transcendence (CT) [6] 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.

Next, we demonstrate this model in a scenario where autonomous agents must make transit choices. Autonomous agents identify with different notions like environmentalism, frugalism, etc., each of which impacts their choices. Eventually, in our transit simulation, agents must choose between public transport or private transport. Using the model of identifying with notions, we can create diverse kinds of autonomous agents identifying with different notions to varying extents and study the collective behaviour which emerges in such populations.

Major contributions presented in this paper are as follows: 1. We present the extended model of Computational Transcendence using which agents can identify with multiple abstract notions to varying extents. 2. We demonstrate an application of this model for the use case of autonomous agents making transit choices. 3. We also analyze the impact of factors like conformity on autonomous agents. The paper is organized as follows: In Section 1, we introduce the topic of autonomous agents identifying with abstract notions. Next, in Section 2, we look at some of the work which has been done in the area of decision-making by autonomous agents and different ways of modelling transit choices. Section 3 presents the detailed formal model of agency using which autonomous agents can identify with abstract notions. Next, we present the experimental details of translating this model to an agent making transit choices in Section 4. Results of the simulations are presented in Section 5. Finally, we conclude with key discussions and conclusions of this work in Section 6.

2 Related Work

In this section, first, we discuss modelling autonomy into artificial agents which affects their decision-making. Next, we look at the importance of modelling a social identity which can be useful to elicit cooperation from agents for bigger causes. Finally, we elaborate on some of the existing work done in the area of modelling transit choices.

2.1 Autonomous Decision Making

The motivation to build artificial agents that can make decisions of their own has been fuelled by the deployment of `AI` in various fields of practical use like healthcare [19, 16, 1], finance [5, 38], autonomous vehicles [26, 24] to name a few. Engineering the ability to make decisions by themselves into agent-based systems is studied under the umbrella of autonomous agency. The AI-based systems built using autonomous agents are expected to make decisions autonomously in real time. Also, these systems must act responsibly and preferably provide reasoning behind their decisions. Thus, modelling autonomous agency is challenging yet increasingly important, given the penetration of AI in society.

Modelling autonomy in artificial agents has been approached from several paradigmatic standpoints. These include Normative models, Adaptive Learning, Rational Choice, and Models of Self [34].

The Normative paradigm focuses on establishing a set of norms or rules that guide the decision-making process of autonomous agents. These norms define the agents' acceptable behaviour and ethical boundaries, enabling them to make decisions that align with predefined principles and values. The Beliefs-Desire-Intention (`BDI`) model [8], Event Condition Action (`ECA`) [13] sequence and the Truth Maintenance Systems (`TMS`) [10] etc. are a few such techniques of modelling agency using the Normative paradigm.

The Adaptive Learning paradigm emphasizes the importance of learning and adaptation in autonomous decision-making. Agents following this paradigm continuously update their decision-making strategies based on the feedback from the environment. Techniques like Reinforcement Learning [35], Swarm Intelligence [22], Ant Colony Optimisation [9], etc. are used to model autonomous agents using this paradigm.

The Rational Choice paradigm is based on the principle of rationality and utility maximization. The Classical Rational Choice Theory has been critiqued for its departure from human behaviour, which often forms the basis of intelligent or autonomous behaviour. A few of the aspects highlighted by these critiques are agents showcasing empathy [30], loss aversion [11], bounded rationality [33], etc. along with maximizing their self-interest.

The Models of Self paradigm involves modelling agents with a sense of self. The idea of a sense of self draws inspiration from identity associations which in turn drive human behaviour. Some models that fall under the umbrella of this paradigm include autonomic computing [14, 23] which models various self-* properties, Autopoiesis [27] which built models from a biological perspective, Cybernetics and Artificial Life [20, 25] which models self-regulatory properties etc.

A recently proposed framework based on the Models of Self paradigm is Computational Transcendence (CT) [6]. CT proposes to model autonomous agents by defining an elastic sense of self. This framework, in turn, allows system designers to define different associations in the environment as part of the identity of individual agents. So far, in practice, only identity associations with other agents in the environment have been modelled using CT. However, it is also important to model identification with abstract notions to design more varied, realistic and responsible autonomous agents. In this paper, we present the extended CT model using which autonomous agents can identify with abstract notions to varying extents and demonstrate diverse behaviour.

Some work has been done on incorporating notions into agents inspired by organisational science, cognitive science, and psychology [29]. Here artefacts are introduced as passive entities that influence the decision-making of the autonomous agents and their applicability in various MAS-based scenarios has been discussed. However, most existing work which incorporates notions in autonomous agents assumes notions as the goals to be achieved. However using the CT framework, notions are the drivers of agents' behaviour rather than a goal to be achieved.

2.2 Social Identity

Humans also identify with abstract notions and strive to uphold them. Social identity is defined as "those aspects of an individual's self-image that derive from the social categories to which he perceives himself as belonging." [36] Also, group identity plays an important role in various social contexts [12]. It is a way to elicit cooperation from the masses for different causes– like climate change, diversity etc.

2.3 Modelling Transit Choices

Psychology has representations of human behaviour using various models, and understanding some of these helps design computational models for autonomous agents. For example, the Theory of Planned Behaviour (TPB) [2] is a widely used theory for modelling human behaviour. It states that the beliefs of an agent, coupled with the subjective norms, form the fundamental basis of an individual's behaviour. It highlights the role of perceived behavioural control in the final behavioural decisions made by the individual.

A recent work [28] in the context of modelling transit choices of humans makes use of TPB to build agents that subsequently model human behaviour in the context of traffic scenarios.

Cui et al. [32] proposed a Multinomial Logit (MNL) based model describing passengers' transit choices using ride-hailing data. This work aimed to suggest how to navigate sustainable ride-hailing choices. They further identified that factors like pick and drop-off locations, travel distance, time, and cost strongly affect passengers' transit choices.

There has been work to estimate distributions of different factors associated with transit, like cost, time, etc. Such studies help model suitable environments by sampling these distributions to simulate the trips of agents [31, 3, 15]. This way of modelling becomes useful, especially when actual data on transit choices is unavailable. Furthermore, even with the availability of transit-related data, these distributions can be perturbed to model a realistic underlying stochastic distribution. We use these latent distributions to construct an environment where different types of transit trips for autonomous agents are simulated.

Agent-based modelling also has been useful in simulating transit choices of people [37, 18, 17]. Further, modelling these traffic flows includes understanding the population density on roads at a given time and the choices people make for their transit decisions. Understanding the interplay of the factors that can influence transit decisions, including beliefs and adaptability of the population, can aid policymakers in planning and designing better transportation solutions. As highlighted by many of the agent-based studies in the context of transit choices, we identify cost, time, comfort (in the form of congestion), environmental awareness (in terms of carbon footprint per head) and social factors (in terms of conformity) as the factors that are relevant for designing the simulation environment. People making transit choices are often individual or local optimizers of their decisions. Thus, devising appropriate strategies means designing systems where local optimizations by individuals lead to a desired globally optimal system state. Such a system state can be quantified in terms of sustainability, low net carbon footprint, etc.

3 Modelling Identification with Notions

In this section, we elaborate on how we model autonomous agents which identify with notions. We make use of Computational Transcendence (CT) [6] for this, a framework used to model autonomous responsible agency. 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. The semantic distance between a and an object in I_a captures how important the object is in shaping the identity of a. The smaller the semantic distance, the greater its influence on the identity of a.
- $\gamma_a \in [0, 1]$ represents the elasticity or transcendence level of the agent a's sense of self.

We extend the CT framework to build agents who can identify with abstract notions. For this, we first define the measurable quantities in an agent's given environment or context as observables. If the context changes, the corresponding observables change. However, the elements in the identity set of an autonomous agent 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. The key problem addressed in this paper is to translate the changes in observables which are measurable from the environment, to the changes in associations with objects in the identity set of an agent. For this, we define a schema for identity objects over the set of observables.

Consider a transcended agent a. Let there be a set of n observables in the given context, $\bar{co} = (co_1, co_2, ...co_n)$. We define the association between an identity object and an observable with a weight. This weight can be any non-negative real number. The greater the weight, the greater the association, while weight 0 means the given observable is irrelevant to the object in the identity set. The relevant objects in the identity set of agent a are the objects which hold a non-zero weight for at least one observable in the given context.

The schema of the identity object is then defined as the normalised weights of the relevant objects in the identity set over all the observables in a given context. For a relevant identity object o_i in I_a of the agent a, the schema of o_i is defined as follows:

$$\bar{s_{o_i}} = (s^i_1, s^i_2, ...s^i_n) \tag{1}$$

$$\Sigma^n_{j=1} s^i_j = 1 \tag{2}$$

The association of an agent with each element in the identity set can be derived from the CT framework [7]. An agent a with a transcendence level γ, having a semantic distance of $d(o_i)$ for an identity object o_i, gives a weight of $\gamma^{d(o_i)}$ to that object. These weights, coupled with the weights defined by the schemas, enable us to compute the final weights an agent assigns to different observables in a context. These final weights can be interpreted as the preferences of the agent.

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 $\bar{s_{o_i}} = (s^i_1, s^i_2, ...s^i_n)$. Then, the preference vector $\bar{p}$ over the observables in that context is defined as follows:

$$\bar{p} = \frac{\begin{pmatrix} \gamma^{d(o_1)} & \gamma^{d(o_2)} & ... & \gamma^{d(o_m)} \end{pmatrix}}{\Sigma^m_{i=1} \gamma^{d(o_i)}} \begin{pmatrix} s^1_1 & s^1_2 & ... & s^1_n \\ s^2_1 & s^2_2 & ... & s^2_n \\ . & . & . & . \\ s^m_1 & s^m_2 & ... & s^m_n \end{pmatrix} \tag{3}$$

$$\bar{p} = \frac{1}{\Sigma^m_{i=1} \gamma^{d(o_i)}} \begin{pmatrix} \gamma^{d(o_1)} & \gamma^{d(o_2)} & ... & \gamma^{d(o_m)} \end{pmatrix} \begin{pmatrix} \bar{s_{o_1}} \\ \bar{s_{o_2}} \\ . \\ \bar{s_{o_m}} \end{pmatrix} \tag{4}$$

Here, the row vector consists of the weights assigned to different objects in the identity set, multiplied by a matrix whose rows correspond to the respective schemas of those objects. This matrix multiplication gives a vector from the row space of the matrix with weights defined by the vector containing $\gamma^{d_{o_i}}$ terms. Thus, the preference vector $\bar{p}$ is a linear combination of schemas corresponding to the objects in the identity set of the transcended agent a.

Given a performable action or choice a_i that results in a set of observables in the given context as $\bar{co}_{a_i}$: $(co^i_1, co^i_2, ..., co^i_n)$, the utility of a_i can be computed as follows, where $\bar{p}$ is expanded as $(p_1, p_2, ..., p_n)$:

$$u(a_i) = (p_1, p_2, ..., p_n) \begin{pmatrix} u(co_1^i) \\ u(co_2^i) \\ . \\ u(co_n^i) \end{pmatrix} \tag{5}$$

$$u(a_i) = \bar{p}.u(\bar{co}_{a_i}) \tag{6}$$

Thus, the final utility of an action or a choice is the dot product of the preference vector $\bar{p}$ and the vector of utilities derived from the observables resulting from an action or a choice. The behaviour of an agent can now be modelled using the standard Markov Decision Process framework. The sense of self of a transcended agent forms the basis of its behaviour.

We now describe how the identity associations are updated. The agent receives some values for the observables as a consequence of its actions. To update its semantic distances, the agent needs to keep track of the average values of observables. These averages can then be used to calculate the utility of averaged observed values. Note that these utilities can also be negative. Let the net utility vector of the observables be: $\bar{nu}_{co} = (nu_{co_1}, nu_{co_2} ..., nu_{co_m})$.

The utility functions map the range of observable values to the interval $[-1, 1]$. This normalisation ensures that despite different ranges of different observables, the updates are consistent for the agent.

We use a threshold to check if the disparity between the net utilities over observables and the schema of an identity object is high enough to induce a semantic distance change. For threshold τ, the semantic distance to an identity object o_i is updated only if:

$$\gamma_a^{d_a^{old}(o_i)}.(\bar{s}_{o_i}.\bar{nu}_{co}) > \tau \tag{7}$$

Let $\bar{s_{o_i}}$ be the schema of an identity object of the transcendence agent a. Then, the updated semantic distance to o_i is defined as follows:

$$d_a^{new}(o_i) = d_a^{old}(o_i) - lr.\gamma_a^{d_a^{old}(o_i)}.(\bar{s}_{o_i}.\bar{nu}_{co}) \tag{8}$$

The term $\bar{s}_{o_i}.\bar{nu}_{co}$ is the dot product of the schema of the identity object o_i with the net utility vector. This term is positive when a schema benefits an agent, and subsequently, the agent reduces its semantic distance to o_i. On the other hand, if the agent gets a negative utility due to a schema, it increases its semantic distance to it. However, this update is only triggered when the term $\bar{s}_{o_i}.\bar{nu}_{co}$ crosses the pre-defined threshold τ. Thresholding ensures that small deviations from the schema do not affect the sense of self of an agent.

Since the utility of the observables is constrained to $[-1, 1]$, the term $\bar{s}_{o_i}.\bar{nu}_{co}$ is always constrained by 1, making the utility calculations consistent. A transcended agent is said to be *stabilised* when it does not change its distance to any of the identity objects in the given environment.

Thus, introducing schemas enables us to translate the utilities computed over the observables to the changes in the identity set of an agent. The objects in the identity set can represent agents in the system, a collective of agents or even notions. There is no standard theory for computing the utilities of these diverse entities and in turn to compute the changes in identity associations. However, the utility of observables is well-studied. Especially theories from behavioural economics, like the prospect theory [11], have mechanisms to estimate the utility of different observables. Thus, introducing schemas bridges the gap between the existing literature and the problem of calculating the utilities of abstract entities like notions.

4 Modelling Transit Choices

Having defined the formal model of autonomous agents that identify with notions, we demonstrate modelling transit choices as a realistic use-case of this model. We translate different objects in the identity set to a composition of observables in a specific context. We encounter the problem of estimating the utility of identity objects and show how defining these objects as schemas over observables helps model identity-based artificial autonomous agents.

As discussed in the model, specific objects in the identity set of an agent become relevant in a specific context, which in turn become drivers for the agent's decision-making. For this, defining the context in which the agents are modelled becomes imperative. We demonstrate an example where the agents in a network need to make decisions about their transit choices. The transit is assumed to be between two fixed points on a periodic basis, such as traveling from home to work on a daily basis. The repetition of acts enables agents to gain feedback from their environment and update their behavior. We have specifically modelled two transit choices, one depicting a private mode of transportation like a taxi service and the other being a public mode of transportation such as a bus.

Various factors in the environment influence the transit choices in humans. For example, we factor cost, time, congestion, and carbon footprint as the relevant contextual observables. Social factors also influence these decisions. We introduce one such factor, conformity, in our model.

In the real world, a population of people is actually a composition of multiple autonomous agents. By modelling the transit choices using artificial autonomous agents, we can simulate different types of populations and study their effects on individuals and the system as a whole. Also, this model can reveal the kind of identity associations that lead to a certain population's behaviour. It can also be used by policymakers and system designers to understand and design interventions in the system in order to achieve particular goals like reducing carbon footprint, improving the efficiency of transit, etc.

4.1 Modelling observables

In the context of making a decision about transit choice, we have identified the following observables:

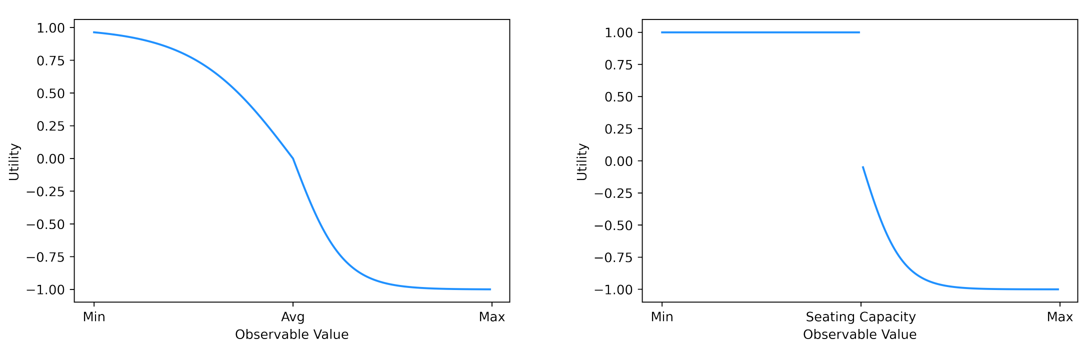
- Cost: The monetary cost incurred by an agent for travelling.
- Time: The time taken by an agent to get from source to destination.
- Congestion: The congestion in terms of occupancy and seating capacity of the mode of transit of the agent.
- Carbon footprint: The carbon footprint per head caused by the chosen mode of transit.

The advantage of dealing with observables is that these can be easily converted into utility. In the model, we use prospect theory to convert the observables - cost, time, congestion, and carbon footprint per head to perceived utilities. All the utility functions are defined such that they map an observable value to the range $[-1, 1]$. This helps in normalizing the observables, since the ranges and units of the observables could vary a lot.

Prospect theory [21] is used to model the utility function for the observables– time, cost, and carbon footprint. It proposes that individuals are risk averse, i.e. the magnitude of utility obtained by a profit of a certain payoff is less than the magnitude of the utility lost by the loss of the same amount of payoff. Figure 1a shows the utility curve that transforms the observable values of cost, time, and carbon footprint to respective utilities. We see that for the least amount of time, cost, or carbon footprint, maximum utility is incurred whereas a maximum value of these observables gives the lowest utility.

Further, the slope of the curve varies on either side of the average. On the positive side of utility, there is a gentle slope, while on the negative side of utility, there is a steeper slope. Thus, there is a kink in the curve at utility zero.

The utility function for the observable congestion is modelled as a discontinuous function as shown in Figure 1b. As long as the occupancy of the vehicle is less than the seating capacity, every passenger gets a seat, and thus, the utility is 1. However, when the occupancy exceeds the seating capacity, discomfort due to congestion is introduced as indicated by the negative utility.



(a) Utility curve to model time, cost, and carbon footprint

(b) Utility curve to model congestion

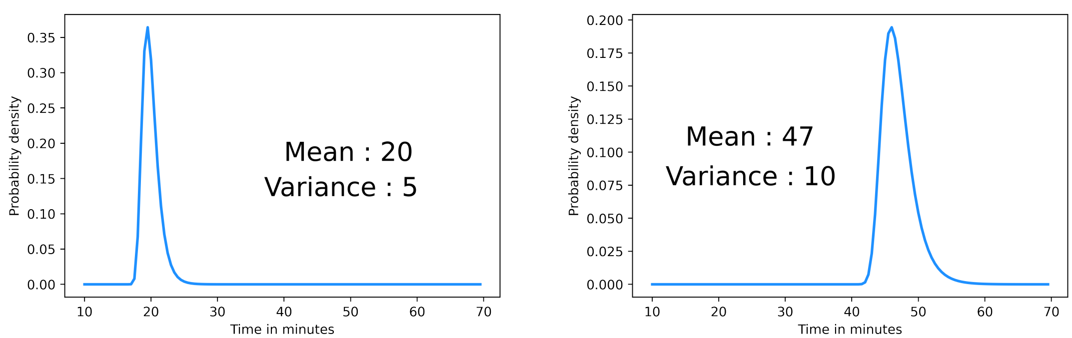
Figure 1: Utility curves modelling observables

4.2 Modelling Choices: Vehicles

Next, we describe how different choices in the context of transit have been modelled. Currently, we have modelled two transit modes, taxi– representing private transport, and bus– representing public transport. However, other modes of transport can also be modelled using this framework. The agent needs to choose between these two transit modes as its commute choice.

For every vehicle on the road (both taxi and bus), the following metrics are tracked:

- *avg cost:* The average cost incurred in transit.
- *avg time:* The average time taken to transit from source to destination.
- *dist:* Distance to be travelled in the current trip.
- *distribution for factors:* Distribution function which can be used to estimate observables for a vehicle.
- *capacity:* The capacity of a vehicle.
- *occupancy:* Current occupancy of a vehicle. It is incremented when agents get assigned to a vehicle.
- *maxOccupancy:* The maximum passenger capacity of the vehicle.
- *emission:* The average carbon emission of the vehicle in grams per kilometre.



(a) Taxi travel time distribution

(b) Bus travel time distribution

Figure 2: Underlying probability distributions to model travel times for transit choices using a right-skewed distribution. The mean and variances for each taxi and bus are highlighted respectively

These parameters can also extended to any mode of transport. For the experiments in this paper, we selected two points in a metro city. Using Google Maps, we estimated the mean and variance of travel time and cost for both modes of transport. Further details of the observables are as follows:

Cost A constant function is used to model the cost of transit for a given vehicle. Specifically, for transit between the two selected points, the average cost of a bus is 20 units and for a taxi is 300 units.

Time A right-skewed distribution, namely the Gumbel distribution is used to model the time taken by the vehicles. The right tail signifies the rare possibility of unusually high travel times, which corresponds to situations like accidents, unforeseen traffic jams, etc. The time of the trip is measured in minutes. Figure 2 shows the underlying Gumbel distributions for taxis and buses respectively.

Occupancy The occupancy of a vehicle is used to calculate the observables - congestion and carbon footprint per head. Occupancy is modelled using two parameters namely - the seating capacity of the vehicle and the maximum occupancy of the vehicle. The actual occupancy of a vehicle can be between zero and maximum occupancy. A Gaussian distribution is used to model the occupancy of a bus, with seating capacity as the mean and one-third of the maximum occupancy as the variance. This represents that on average, the occupancy of the bus is around its seating capacity. Sampling from the Gaussian distribution returns a real number which is adjusted to the range of [0, max occupancy], and then rounded off to the nearest whole number to return the occupancy of the bus. In the case of taxis, a discrete probability distribution is used to model the occupancy. These distributions are plotted in Figure 3 for both taxis and buses. Based on occupancy, the observables- congestion and carbon footprint per head for each vehicle are defined as follows:

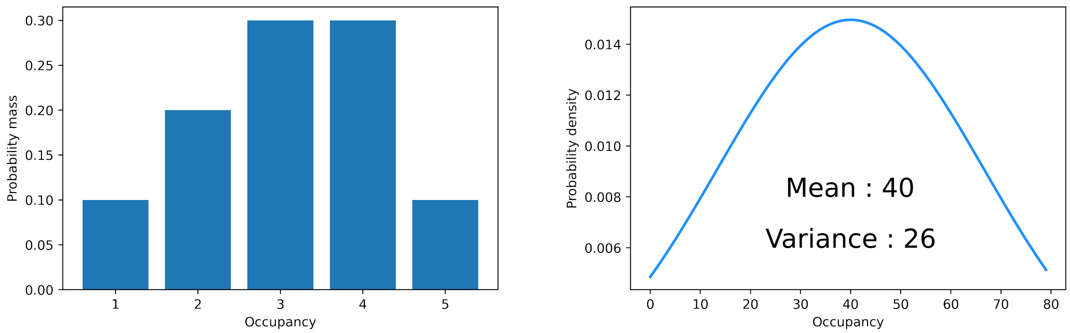


Figure 3: Underlying probability distributions that model occupancy for transit choices of taxi and bus respectively

Congestion

$$\text{Congestion} = \frac{\text{Occupancy}}{\text{Seating Capacity}} \tag{9}$$

The seating capacity for a taxi is 4, while that for a bus is 40.

Carbon foot print

$$\text{Carbon foot print} = \frac{\text{Total emission}}{\text{Occupancy}} \tag{10}$$

The total emission is calculated based on the estimates of the carbon emissions of vehicles. For taxis, it is estimated at 40 grams of carbon per kilometre, and for buses, it is estimated at 200 grams of carbon per kilometre.

4.3 Modelling an Agent

We now introduce a set of non-exhaustive relevant notions that could shape an agent's identity in the context of making transit choices.

Frugalism: Being frugal refers to the quality of being careful when using resources, especially money. In the context of transit choice, this notion constitutes a behavioural choice that incurs lower monetary costs.

Idealism: Idealism as a notion refers to an ideology where agents are concerned about some ideal or utopian goal, and the practicalities of a decision come next. In the case of transit choices, an ideal goal can be considered as the sustainability of the environment, concern about climate change, etc.

Individualism: Individualism as a notion puts emphasis on the agent's freedom, and individual identity. In the case of transit choices, individualism emphasises the idea of individual comfort and individual time incurred in commute.

Pragmatism: Pragmatism as a notion is about being practical. Pragmatism is concerned with all practically measurable quantities in the given context.

As discussed earlier, in mapping notions in the agent's identity to observables in the environment, we need a schema. Table 1 represents the heuristic schema we have come up with to reconcile the definition of different notions with the observables in the context of making transit choices.

Observable / Notion	Cost	Time	Congestion	C Footprint
Frugalism	$\frac{3}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	0
Idealism	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{7}{10}$
Individualism	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}$	0
Pragmatism	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0

Table 1: Schemas of notions over observables

Usually, agents representing a single notion are modeled in multi-agent setups. However, our extended CT framework can be effectively used in representing the identity of each agent as a combination of multiple notions.

Agents are simulated in a network and based on their identity associations, they make transit choices.

They estimate the values of each observable for each transit choice probabilistically based on the perceived utility of each transit choice over time. The perceived utility is computed based on the average value of observables. The sense of self of the agent is represented in terms of identity associations with the notions. Also, agents update their identity associations on a slower time scale based on the distance update logic discussed previously.

We also model conformity among the agents based on the network structure. Conformity introduces the social aspect to the system and it models the way in which agents are affected by their neighbours' transit choices. 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:

$$u(c_i) += cf * frac\_neigh_{c_i} \tag{11}$$

Here, cf is the *conformity factor* which is a measure of the extent of the conforming tendency of the agents in the network. cf can be in the range of $[0, 1]$, with 0 representing agents driven by their individual identities and 1 representing agents who are influenced by their neighbourhood as much as their individual preferences.

5 Experiments and Results

This section presents the results and inferences drawn from different simulation-based experiments. 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.

5.1 Baseline

We first simulate a baseline case. Since conformity is ubiquitous in real-world human decision networks, we assume a conformity factor of 0.2 in our baseline. The semantic distances to different notions for each agent are randomly initialised by uniform sampling between $d \in [0, 4]$,. We run the simulation on an Erdős–Rényi graph with 500 agents. Each agent has a transcendence level $\gamma = 0.8$. With this initialisation of γ and d the value γ^d lies approximately between 0.5 and 1, ensuring each notion has a significant impact on the agent's behaviour at the beginning. Semantic distance updates are performed after every epoch, which comprises 10 trips.

We run the simulation till the agents in the network stabilise. In the stabilised network, we note that only around 33.5% of the population chooses public transport. Implying that in our baseline case, agents are inclined towards choosing private transport.

Next, we examined the average variation in semantic distances of agents in the network for different notions in this context. Figure 4: 'Whole Population' - shows the trends of semantic distance to each notion averaged over the agents with the progress of the trips.

We investigated the average semantic distances of the two sub-populations based on their choice for more insights on how notions drive decisions. Figures 4 : 'Taxi Population,"Bus Population' - aggregate results for sub-populations that chose taxi and bus, respectively. The average semantic distances of these trip-wise sub-populations are plotted. Note that these plots are plotted with every trip(not epoch) because of the stochastic nature of an agent's decision; the sub-population itself might change every trip, though semantic distance update only happens after an epoch. The following inferences can be made from these plots:

- After stabilisation, in the sub-population of agents that choose taxis, individualism is the dominant notion (having the lowest average semantic distance). While in the case of buses, the dominant notion is frugalism.
- Notions like frugalism and individualism are strong indicators of sub-population preferences. By strong indicators, we mean the absolute difference between the average semantic distance to a notion between the sub-populations is high. For strong indicators, this difference is greater than or equal to one.
- Notions like idealism and pragmatism are weak indicators of preferences. In this case, the absolute difference between the average semantic distance to notions between the sub-populations is close to zero.

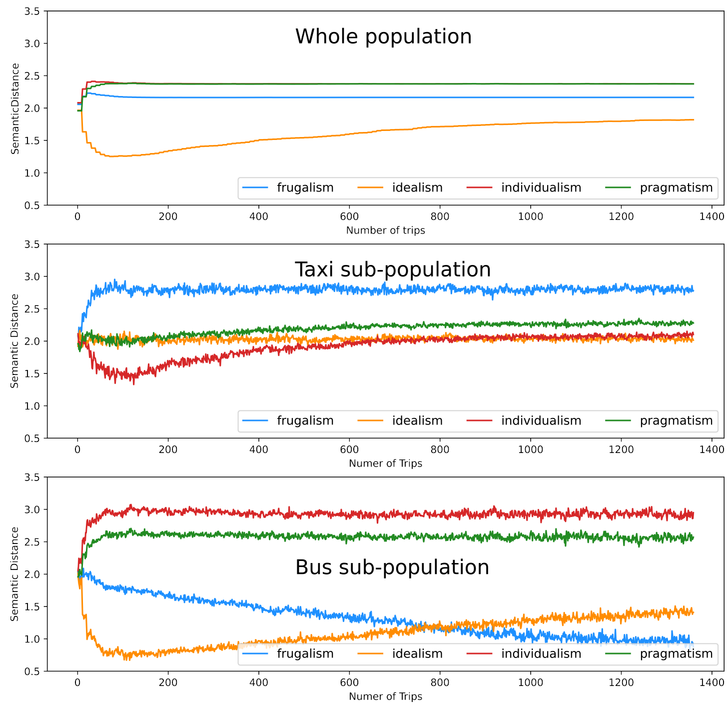


Figure 4: Varying of average semantic distances of the total population and sub-populations, choosing a particular transit choice, to different notions over the epochs

5.2 Varying Semantic Distances

In this experiment, we study the impact of the initialisation of semantic distances on the settled transit choice of the population. To understand the impact of each notion on the settled transit choice, systematically, the average distance to a specific notion is reduced. This is done by sampling the semantic distance for a specific notion uniformly from the range of $[0, 2]$ instead of $[0, 4]$.

Table 2 summarises the four configurations where each of the four notions is given an average lower semantic distance, implying closer association to that notion in the population. The blue and orange bars represent the initial and final average semantic distance to each notion.

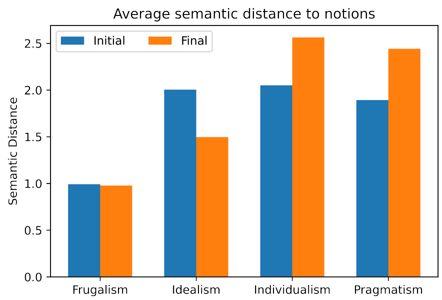
We observe that the initial distributions affect the settled population's choice. Thus, there does not exist one unique equilibrium in this complex system; rather, depending on the initial beliefs of the agents in the system, the resultant behavior of the stabilized population emerges. It also implies that in order to achieve a specific system state, the interventions must depend on the starting state of the system. A one-size-fits-all solution for policy makers.

The extended CT framework gives the capability to capture this diversity in agent behaviour. In the next section, we analyse the effect of conformity in these networks.

5.3 Varying the Extent of Conformity

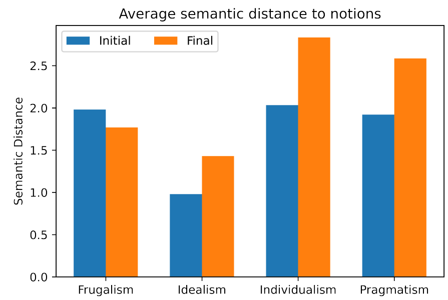
The edges of the network in which we carry out our simulation indicate an interaction between agents. The neighbourhood of an agent a consists of all the other agents in the network whose transit choice influences the decision of a. In this formulation, we observe the effect of the strength of conformity between every agent and its neighbourhood. This effect is captured by scaling the utility that an agent derives through conforming, which is proportional to the ratio of agents in its neighbourhood with the same decision or transit choice during a trip.

Frugalism



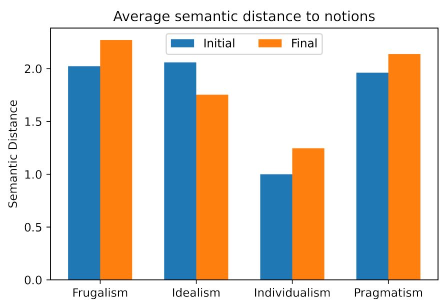
Public transit proportion: 0.642

Idealism



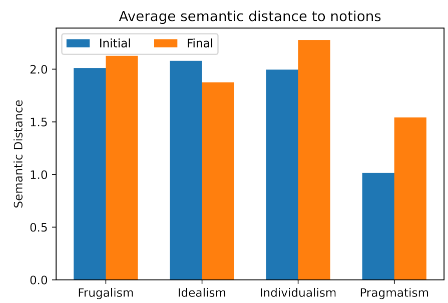
Public transit proportion: 0.476

Individualism



Public transit proportion: 0.214

Pragmatism



Public transit proportion: 0.336

Table 2: Initialization of average semantic distance to a particular notion reduced to see its effect on the stabilized network.

The heatmap in Figure 5 shows the relation between the initial semantic distance distribution and the extent of conformity among the agents in the network. The following observations can be made from this plot:

The extent of conformity influences the polarity of the population's transit choice. We observe from Figure 5 that with an increase in the extent of conformity in the network, the polarity or the strength of transit choices of the population increases. When there is no conformity among the agents, every agent makes choices independently based on their identity associations with different notions. However, with high conformity among the agents, both an individual's identity and conformity with the neighbourhood affect their decision-making.

With the increase in conformity factor, the overall decision of the population of agents initialised towards a particular notion tends to become polarised. This phenomenon can be understood as a consequence of the relative dominance of utility derived from conforming over the utility derived from an individual's identity associations. Hence, weak preferences at a population level develop into an agent's identity associations, bringing out a strong preference collectively, and this loop keeps reinforcing an agent. However, when the conformity factor is lower, such polarization cannot be observed because the initial weak preferences at the population level remain confined to the individual level.

The notion of conformity suggests that an agent must conform with its neighbours, which leads to a reduction in the diversity of behavioural choices of the population. Thus, modelling autonomous agents using the extended `CT` framework resurfaces the impact of conformity in a network of agents.

Strong and Weak Indicators of Transit Choice Notions like Frugalism and Individualism are strong indicators for behavioural choices made by the agents, while Idealism and Pragmatism are weak indicators. We made this observation in our first experiment. However, this also becomes apparent through this experiment. From Figure 5, we observe that in scenarios where the population is initialised towards a notion that is a strong indicator of a behavioural choice, the overall choice of the settled population becomes more polarised at relatively smaller conformity factors as compared to the weak indicators.

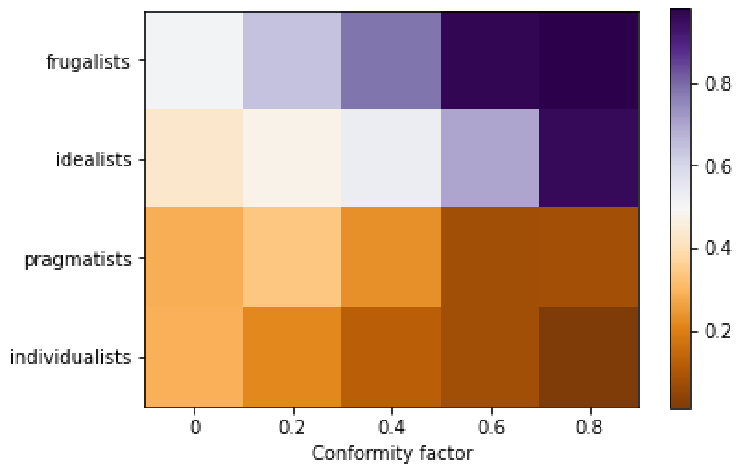


Figure 5: Heatmap of the effect of conformity and initialization of semantic distance to notions on the proportion of stabilized population choosing public transport

6 Conclusions

Our world is a composition of multiple autonomous agents. By modelling the transit choices using artificial autonomous agents, we can simulate different types of populations and study the effects of social factors and other factors that influence the identity of individuals and impact the system as a whole. This model can reveal the kind of identity associations that lead to 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 particular goals like reducing carbon footprint, improving the efficiency of transit, etc.

The extended CT model helps to address the problem by modelling an identity in autonomous agents. Using this model, we build agents that identify with different abstract notions to different extents. These abstract notions connect to real-world observables using our proposed mechanism 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. This in turn helps to build autonomous agents having an identity that can dynamically adapt.

We use this to model autonomous agents making transit choices. The factors affecting these transit choices have been identified in terms of direct observables like time, cost, congestion, environment, etc. Also, social factors play an essential role in shaping their behaviour. The extended CT model effectively models agent-level behaviour in terms of direct observables and also allows the introduction of network-level social factors like conformity. We believe this framework is extensible to account for various individual and social factors relevant in a given context. Also, it can be used to understand the effects of different policy interventions and help the system designers develop effective interventions leading to sustainable population choices across diverse applications.

Acknowledgement

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] Michael D Abràmoff, Philip T Lavin, Michele Birch, Nilay Shah, and James C Folk. Pivotal trial of an autonomous ai-based diagnostic system for detection of diabetic retinopathy in primary care offices. *NPJ digital medicine*, 1 (1):39, 2018.

[2] Icek Ajzen. The theory of planned behavior. *Organizational behavior and human decision processes*, 50(2): 179–211, 1991.

[3] Beda Büchel and Francesco Corman. Review on statistical modeling of travel time variability for road-based public transport. *Frontiers in Built Environment*, 6:70, 2020.

[4] Ralph Buehler and John Pucher. Cycling to sustainability in amsterdam. *Sustain*, 2009.

[5] Longbing Cao. Ai in finance: challenges, techniques, and opportunities. *ACM Computing Surveys (CSUR)*, 55(3): 1–38, 2022. doi:10.1145/3502289.

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

[7] Jayati Deshmukh, Nikitha Adivi, and Srinath Srinivasa. Modeling application scenarios for responsible autonomy using computational transcendence. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pages 2496–2498, 2023.

[8] Frank Dignum, David Morley, Elizabeth A Sonenberg, and Lawrence Cavedon. Towards socially sophisticated bdi agents. In *Proceedings fourth international conference on multiagent systems*, pages 111–118. IEEE, 2000.

[9] Marco Dorigo and Gianni Di Caro. Ant colony optimization: a new meta-heuristic. In *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, volume 2, pages 1470–1477. IEEE, 1999.

[10] Jon Doyle. A truth maintenance system. *Artificial intelligence*, 12(3):231–272, 1979.

[11] Kimberley D Edwards. Prospect theory: A literature review. *International review of financial analysis*, 5(1): 19–38, 1996.

[12] Naomi Ellemers, Russell Spears, and Bertjan Doosje. Self and social identity. *Annual review of psychology*, 53(1): 161–186, 2002.

[13] S. Franklin and A. Graesser. Is it an agent, or just a program?: A taxonomy for autonomous agents. *In International workshop on agent theories, architectures, and languages*, page 2135, 1996.

[14] Alan G Ganek and Thomas A Corbi. The dawning of the autonomic computing era. *IBM systems Journal*, 42(1): 5–18, 2003.

[15] Younes Guessous, Maurice Aron, Neila Bhouri, and Simon Cohen. Estimating travel time distribution under different traffic conditions. *Transportation Research Procedia*, 3:339–348, 2014.

[16] Varun Gulshan, Lily Peng, Marc Coram, Martin C Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Kasumi Widner, Tom Madams, Jorge Cuadros, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22):2402–2410, 2016.

[17] Banafsheh Hajinasab, Paul Davidsson, Jan A Persson, and Johan Holmgren. Towards an agent-based model of passenger transportation. In *Multi-Agent Based Simulation XVI: International Workshop, MABS 2015, Istanbul, Turkey, May 5, 2015, Revised Selected Papers 16*, pages 132–145. Springer, 2016.

[18] Jiangyan Huang, Youkai Cui, Lele Zhang, Weiping Tong, Yunyang Shi, and Zhiyuan Liu. An overview of agent-based models for transport simulation and analysis. *Journal of Advanced Transportation*, 2022, 2022.

[19] Fei Jiang, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. Artificial intelligence in healthcare: past, present and future. *Stroke and vascular neurology*, 2(4), 2017.

[20] John Johnston, Morris Moscovitch, Carlo Umiltà, et al. *The allure of machinic life: Cybernetics, artificial life, and the new AI*. MIT Press, 2008.

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

[22] James Kennedy. *Swarm intelligence*. Springer, 2006.

[23] Jeffrey O Kephart and David M Chess. The vision of autonomic computing. *Computer*, 36(1):41–50, 2003.

[24] Hamid Khayyam, Bahman Javadi, Mahdi Jalili, and Reza N Jazar. Artificial intelligence and internet of things for autonomous vehicles. *Nonlinear Approaches in Engineering Applications: Automotive Applications of Engineering Problems*, pages 39–68, 2020.

[25] Maciej Komosinski and Andrew Adamatzky. *Artificial life models in software*. Springer Science & Business Media, 2009.

[26] Yifang Ma, Zhenyu Wang, Hong Yang, and Lin Yang. Artificial intelligence applications in the development of autonomous vehicles: A survey. *IEEE/CAA Journal of Automatica Sinica*, 7(2):315–329, 2020.

[27] Humberto R Maturana and Francisco J Varela. *Autopoiesis and cognition: The realization of the living*, volume 42. Springer Science & Business Media, 1991.

[28] Ali Najmi, Travis Waller, Mehrdad Memarpour, Divya Nair, and Taha H. Rashidi. A human behaviour model and its implications in the transport context. *Transportation Research Interdisciplinary Perspectives*, 18:100800, 2023. ISSN 2590-1982. doi:https://doi.org/10.1016/j.trip.2023.100800. URL `https://www.sciencedirect.com/science/article/pii/S2590198223000477`.

[29] Andrea Omicini, Alessandro Ricci, and Mirko Viroli. Artifacts in the a&a meta-model for multi-agent systems. *Autonomous agents and multi-agent systems*, 17:432–456, 2008.

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

[31] Chaoyang Shi, Bi Yu Chen, and Qingquan Li. Estimation of travel time distributions in urban road networks using low-frequency floating car data. *ISPRS International Journal of Geo-Information*, 6(8):253, 2017.

[32] Yuchao Cui; Hongzhi Guan; Zhengtao Qin; Yang Si. Research on the choice behavior of different types of ride-hailing services. *20th COTA International Conference of Transportation*, pages 3807–3819, 2020.

[33] Herbert A Simon. Bounded rationality. *Utility and probability*, pages 15–18, 1990.

[34] Srinath Srinivasa and Jayati Deshmukh. Paradigms of computational agency. *CoRR*, abs/2112.05575, 2021. URL `https://arxiv.org/abs/2112.05575`.

[35] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

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

[37] Kay W Axhausen, Andreas Horni, and Kai Nagel. *The multi-agent transport simulation MATSim*. Ubiquity Press, 2016.

[38] Xiao-lin Zheng, Meng-ying Zhu, Qi-bing Li, Chao-chao Chen, and Yan-chao Tan. Finbrain: when finance meets ai 2.0. *Frontiers of Information Technology & Electronic Engineering*, 20(7):914–924, 2019.