

## A Socio-Psychological Approach To Simulate Trust and Reputation in Modal Choices

Khoa Nguyen<sup>1</sup> and René Schumann<sup>1</sup>

University of Applied Sciences Western Switzerland (HES-SO Valais Wallis)  
SiLab, Rue de Technpole 3, 3960 Sierre, Switzerland  
{khoa.nguyen, rene.schumann}@hevs.ch

**Abstract.** Trust and reputation are currently being researched broadly in multiple disciplines and often considered as main drivers for human's actions in many different scenarios. In the multiagent based simulation community, there are still concerns about qualifying and modelling them with sufficient details and adequateness in the context of decision-making. Besides, the diversity of application domains requires a method to combine trust and reputation with other determinants to provide a more complete picture for the deliberating process in complex systems. This paper presents a novel solution by applying subjective logic in conjunction with a modelling framework that utilises Triandis' Theory of Interpersonal Behaviour to simulate the modal choices of human individuals as an example. It uses the concept of opinion as a metric to measure the belief of an agent about the consequence(s) of an action, which can be updated through feedback. In addition, its consensus rule allows us to combine relevant opinions of the neighbours to evaluate the reputation of the target. By performing an experiment set up in the mobility domain, we demonstrate the framework ability to capture the ground truth of a service's reputation at different simulation scales and highlight the effects of these concepts on the figure of yearly rail-kilometres travelled.

**Keywords:** Trust and Reputation · Social simulation · Agent-based modelling.

### 1 Introduction

Creating autonomous agents capable of exhibiting human-like behaviours under uncertainty of complex environment is becoming increasingly relevant in the simulation research community [12, 21]. One of the most popular methodologies for this purpose is to utilise trust and reputation to give an idea of the confidence one can have on the quality of potential cooperation or recommended systems [17]. However, one of the main concern for using these concepts is transparency [11], which includes human interpretability and explanations of ontologies in different application domains. Several models of trust and reputation have been proposed in the literature, such as those suggested in [10] and [17]. There is still, however, a lack of research efforts that emphasize on understanding the roles of them and their relationships with other determinants in human decision-making.

We aim to provide a framework that can address the matter of limited expressibility and linking highly abstract concepts with empirical data. It employs Triandis' Theory of Interpersonal Behaviour (TIB) [35] to provide a meaningful set of determinants that contribute to decision-making in socio-psychological research. TIB is a tri-level model that allows users to have the freedom of separating and highlighting the impacts of different psychological concepts to individuals and their society. To represent trust and reputation, we incorporate subjective logic to measure personal beliefs about available modal choices presented to the agent. Using feedback(s) from the environment, an agent has the capability to evolve its decision-making process over time depending on personal experiences and opinions from neighbours in its network.

The structure of this paper is as followed: We first consider some of the related projects that model trust and reputation at the individual level (Section 2). It is followed by a specification of our architecture, which contains the functions to calculate subjective probability and utility for each available options. Next, a case study of the implementation platform - BedDeM - is described in Section 4. Its main purpose is reproducing the collective ground truth of reputation for the rail services in both regional and national level and evaluating the impacts of it on the number of kilometres travelled. We then conclude our work and suggest further researching directions in Section 5.

## 2 Related Work

The state-of-the-art in agent's modal choice has been previously discussed in our papers [24, 25]. This study focuses on the individual-level trust and reputation models for agent-based simulation purposes. One of the common approaches is for the agent to interact and study their behaviour patterns over a number of encounters, i.e *learning models*. In this case, the outcome of direct interaction is the measurement of trust. Another method involves the agent asking its network about their perceptions of the target(s), i.e *reputation models*. The final category is *socio-cognitive models*, which concerns with forming coherent beliefs about different characteristics of the partner(s) and reasoning about these beliefs to decide how much trust should one put in them.

### 2.1 Learning models

The most common example to demonstrate the evolution of trust or cooperation over multiple interactions is Axelrod's tournaments around Prisoner's Dilemma [3]. By allowing agents to adapt and evolve their relationship using the cooperation strategy over multiple interactions, Wu et al. [38] showed that trust can actually emerge between them. Their model, however, does not take into account the fact that there might be some utility loss in short-term evaluation. Sen et al. [31] demonstrated how reciprocity can emerge when the agent learns to predict the future benefits of cooperation. In case the number of interactions does

not allow trust to be built, Mukhejee et al. [23] showed how trust can be acquired through mutually learning. In addition, Birk’s learning method [6] allows agents to cooperate in a non-competitive environment. However, all of the models above assume complete information for the multi-agent learning algorithms to work and their results can only be obtained under very strict assumptions.

To become more realistic, the agent needs a way to evaluate how the target’s action affect its goals using the data gathered through their interactions. Thus, many trust metrics have been derived. In the context of trading networks, Witkowski et al. [37] proposed a model; in which the trust for a partner is calculated through equations based on their past interactions and its update vary depending on the type of agent. Other models consider the quality of agents to be a bi-state value (good/bad), such as those of Scillo et al. [30], Mui et al. [22] and Sen and Sajja [32]. For open distributed systems that contain a richer set of outcomes, the REGRET model, which is developed by Sabater et al. [29], attributes some fuzziness to the notion of performance. Its evaluation of trust is not only based on an agent’s direct perception of its partner’s reliability but also their behaviours with others in the system. However, this can imply a number of problems if the information perceived by the agent is wrong or incomplete.

## 2.2 Reputation models

In this subsection, we consider the reputation, which is mainly derived from an aggregation of opinions from members of a community about one individual. Yu et al. [40] use the concept of referrals, which are represented by pointers to other sources of information, similar to links on a Website. In this context, they show how the agent can explore a network using referrals gathered from their neighbours to gradually build up a model of its network. The representation of social networks was extended by the work of Schillo et al. [30]. They annotate the nodes with characteristics, such as honesty and altruism, which can then be used to deduce the trustworthiness of witnesses when calculating the reputation of potential partners. Higher-level concepts (e.g. group and neighbours) are utilised in [29, 40] to related the closeness with the reliability of the witnesses. Nevertheless, all of the models above share the assumption that witness share information freely (i.e. without any cost or personal agenda).

Another popular methodology is aggregating ratings, whose applications come from mostly online communities. A typical example is eBay peer-to-peer rating system [33]. It can however, be unreliable when the users do not return any ratings or sellers try to manipulate the result to build their reputation. The model developed in [41] to deal with the absence of data using referrals and Dempster Shafter’s Theory of evidence [39]. It treats the lack of belief as to the state of uncertainty and allows the combination of beliefs obtained from various sources. The models in [30, 32] extended this work and demonstrated how agents cope with lying witness through learning rather than attributing subjective probability. Another method from Sabater et al. (i.e. REGRET system) [29] adopts the view from sociology to allow realism in term of ontological dimension. The reputation value in which is a weighted sum of subjective impressions from

direct interactions, its group’s impression of the target, its group’s impression on the target’s group and the agent’s impression on the opponent’s group. Despite that, REGRET does not handle the problem of a lying agent, i.e. ratings are not obtained in competitive environments but a more cooperative setup.

### 2.3 Socio-cognitive models

The methods from previous sections are based on an assessment of the outcomes of interactions. At the same time, it is also important to consider the subjective perception of an individual on the indirect interaction to enable a more comprehensive analysis of characteristics of the potential partner [16]. The line of works in this respect was initialised by the work of Castelfranchi and Falcone [8, 9, 14], especially for the Belief-Desire-Intention agent architecture. It is strongly motivated by human behaviours, which are not always rational. Another strategy of this type is from Brainov and Sandholm [7], which supports the need to model an opponent’s trust with a rational approach.

Socio-cognitive modelling research is still in its early stage. Thus, these models mentioned above are often criticised for the lack of grounding in the rational mechanisms [28] which learning and reputation models provide. They could also exploit the assessment performed by the other model types by taking into account learning strategy over multiple interactions, the reputation of potential partners and their motivations and abilities regarding the interactions. However, it can be computationally expensive to allow social agents to reason about all the factors affecting their trust in others. Moreover, these agents are limited in their capacity to gather information from various sources that populate their environment [28].

## 3 Agent decision-making architecture

In this paper, we propose a socio-cognitive agent architecture that can capture the roles of trust and reputation in decision-making. The first subsection illustrates how these concepts are being modeled using the subjective logic methodology. It is followed by the description of the decision-making mechanism, which implements the Triandis’ Theory of Interpersonal Behavior to derive utility values for a set of alternatives options. Finally, we provide an overview of our agent architecture framework.

### 3.1 Modelling trust and reputation

A fundamental aspect of the human condition is that nobody can determine with absolute certainty whether a preposition about the world is true or false. Thus, whenever a statement is assessed, it should be done under the view of an individual and not to be represented in a general and objective belief [19]. Reviews from [5] and [18] provide good examples of how standard logic and probabilistic

logic are designed for an idealised world, where these important aspects are often not considered and conclusions have to be drawn from insufficient evidence.

In this study, we follow framework proposed by Jøsang et al. [20] on subjective logic. It presents a specific calculus that uses a metric called *opinion* to express our subjective beliefs about the world. An opinion, denoted as  $\omega_x^A = (b, d, u, a)$ , indicates the party  $A$ 's belief in the statement  $x$ . In this case,  $b, d$  and  $u$  represents *belief*, *disbelief* and *uncertainty* respectively, where  $b + d + u = 1$  and  $b, d, u \in [0, 1]$ . All set of possible opinions can be mapped into a point inside an equal-sided triangle in Figure 1.

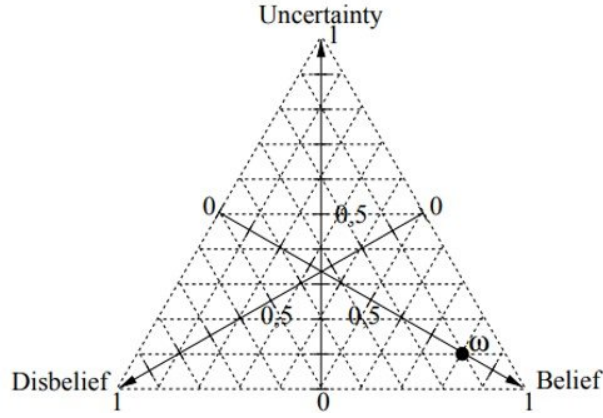


Fig. 1: The opinion-space trust model / opinion triangle [20]

The base rate parameter,  $a \in [0, 1]$ , determines how uncertainty shall contribute to the probability expectation value:

$$E(\omega_x^A) = b + au \quad (1)$$

In the binary event space (i.e. where there are only two possible outcomes - success or failure), subjective logic allows us to build an opinion from a set of evidences about  $x$  using the following equation:

$$\omega_x = \begin{cases} b_x = \frac{r}{r+s+W} \\ d_x = \frac{s}{r+s+W} \\ u_x = \frac{W}{r+s+W} \\ a_x = \text{base rate } x \end{cases} \quad (2)$$

where  $r$  is the number of positive evidences about  $x$ ,  $s$  is the number of negative evidences about  $x$ ,  $W$  is the non-informative prior weight, also called a unit of evidence, normally set to 2 and the default value of base rate  $a$  is usually set at  $1/2$ . In the case of no prior experience with the target, agent  $A$ 's opinion

of  $x$  is set as  $\omega_x^A = (0, 0, 1, 1/2)$ . Therefore, its probability expectation value is  $E(\omega_x^A) = 1/2$ .

We model reputation in this case using the consensus rule of independent opinions A and B:

$$\omega_x^{AB} = \omega_x^A \oplus \omega_x^B = \begin{cases} b_x^{AB} = \frac{b_x^A u_x^B + b_x^B u_x^A}{k} \\ d_x^{AB} = \frac{d_x^A u_x^B + d_x^B u_x^A}{k} \\ u_x^{AB} = \frac{u_x^A u_x^B}{k} \\ a_x^{AB} = \frac{a_x^A u_x^B + a_x^B u_x^A - (a_x^A + a_x^B) u_x^A u_x^B}{u_x^A + u_x^B - W u_x^A u_x^B} \end{cases} \quad (3)$$

where  $k = u_x^A + u_x^B - u_x^A u_x^B$ .

### 3.2 Utility function

In order to create a system that can mimic the function of a human network, the first question to address is how to model our behaviours. In psychology, different theories in the school of cognitive model describe this process, e.g. Ajzen and Fishbein's Theory of Reasoned Action [15] and Ajzen's (1991) Theory of Planned Behaviour [1], etc. We decide to implement Triandis' Theory of Interpersonal Behaviour (TIB) [35] due to its inclusion of a more comprehensive set of determinants and the ability to combine these to form a complete picture of human decision-making. This set is also flexible enough to reflect other behaviour theories by exchanging the determinants and/or assigning weights to mark their contribution to the agent's reasoning process. Figure 2 shows a tree-structure representation of the main determinants in TIB.

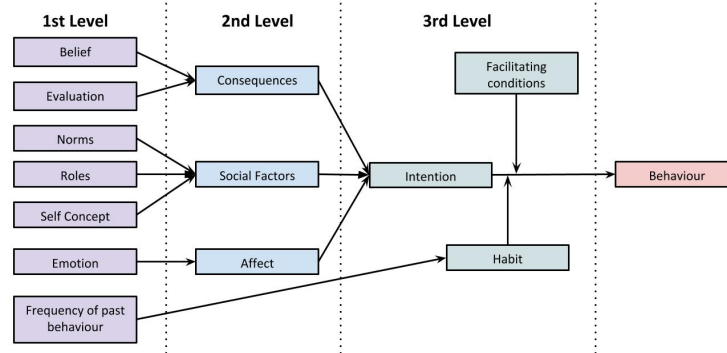


Fig. 2: Triandis' tri-level model [35]

Triandis suggests that one of the main factor to determine intention of an act is the value of perceived consequences, C, depending on the sum of the products of the subjective probability that a particular consequence will follow

a behaviour ( $P_c$ ) and the value of (or affect attached to) that consequence ( $V_c$ ) (see page 16 [35]). Thus, the equation for expected utility of C is as followed:

$$EU_c = \sum_{i=1}^n (P_{c_i} V_{c_i}) \quad (4)$$

where  $n$  is the number of consequences that a subject perceives as likely to follow a particular behaviour. The  $P_{c_i}$  value can be derived from the Equation 1, i.e.  $P_{c_i} = E(\omega_{c_i})$ .

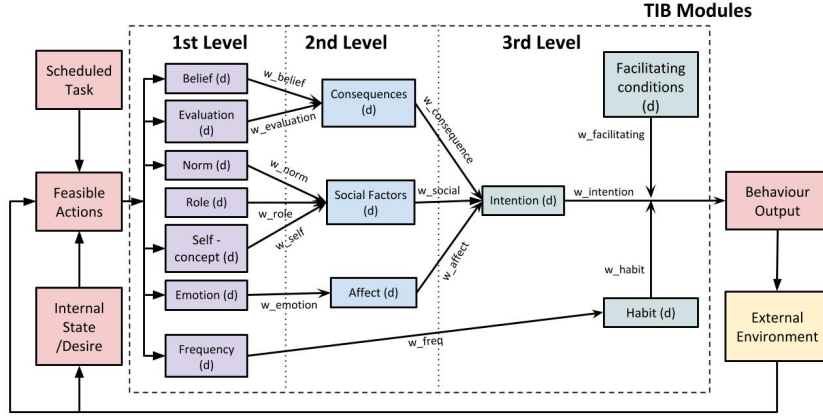


Fig. 3: Agent decision-making mechanism with TIB Module

A full decision-making cycle with TIB Module is illustrated in Figure 3. An agent first selects an isolated decision-making task from the list that is sequentially executed. Its personal desire/goal is then combined with means provided by the external environment to generate a set of possible propositions/options. For all determinants ( $d$ ), each option ( $opt$ ) is given a utility value which comes from comparing its property with other's ( $U_d(opt)$ ). In the first level, this value can be in the form of a real numerical system (for determinants such as price or time) or ranking function (for determinants such as emotion). Either of which can be calculated from empirical data (e.g. census, survey) or calibrated with expert's knowledge and stakeholders' assessment. The results for these determinants are then normalized and multiplied with an associated weight (called  $w_d$ ); the sum of which becomes the referenced value for the option in the next level. This process is captured in the following equation:

$$EU_d(opt) = \sum_{a=1}^A (EU_a(opt) * w_a / (\sum_{o=1}^O EU_a(o))) \quad (5)$$

where  $EU_d(opt)$  is the utility value of an option ( $opt$ ) at determinant  $d$ .  $A$  is the set of all ancestors of  $d$  (i.e. determinants connects with  $d$  in the previous

level).  $O$  is the set of all available options.  $w(a)$  is the weight of ancestor  $a$ . In this case, the weight represents the importance of a decision-making determinant compare to others at the same level and emphasizes on the heterogeneity of individuals. It also allows the modeler to express a certain theory by cutting determinants (i.e. setting their values to 0) that are not relevant to a case study. The combination process then continues until it reaches the behavior output list; the utility value of which can be translated to the probabilities that an agent will perform that option. If the agent is assumed to be deterministic, it picks the option that is correlated to the highest or lowest utility depending on modeler's assumptions.

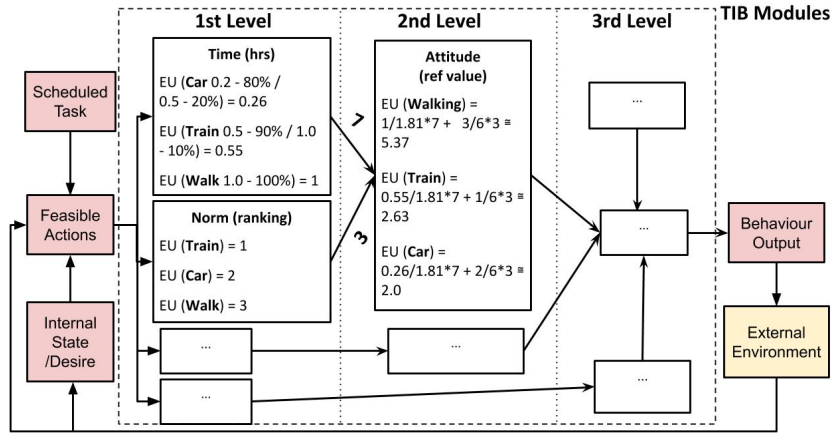
In our framework, the utility function of determinant *Consequences* is an exception case which follows Equation 4. In addition, the expected value of determinant *Norm* can be derived from the probability expectation value of the collective opinion formed by the consensus rule (see Equation 1 and Equation 3). In Section 4, we focus on a simplified binary event space, i.e. an action has two outcomes - success or failure. More complex scenarios could be considered in the future by extending the result space to multiple dimensions (e.g. time, cost, satisfaction).

A running example in mobility domain can be seen in Figure 4.  $EU$  in this case is a cost function, i.e. option that has smaller value is referred. We assume that an agent has access to 3 options: *walking*, using *car* or taking *train*. It expects a *car* journey would take around 0.2 hours for good traffic, which is believed to be 80% chance. A late *car* drive would take up to 0.5 hour. Using subjective logic Equation 4, we have  $EU_{Time}(car) = 0.2 * 80\% + 0.5 * 20\% \approx 0.26$ . Similarly,  $EU_{Time}(train) \approx 0.55$  and  $EU_{Time}(walking) \approx 1$ . Their total value,  $\sum EU_{Time}$ , is 1.81. For non-measurable value such as *Norm*, the agent uses the concept of reputation to rank the options:  $EU_{Norm}(train) = 1$ ,  $EU_{Norm}(car) = 2$ ,  $EU_{Norm}(walking) = 3$  (best to worst); the sum of which is 6. If  $w_{Time}$  and  $w_{Norm}$  are 7 and 3 respectively, the new expected value in next level ( $EU_{Attitude}$ ) of *walking* would be  $1/1.81*7 + 3/6*3 \approx 5.37$ , *train* would be  $0.55/1.81*7 + 1/6*3 \approx 2.63$  and *car* would be  $0.26/1.81*7 + 2/6*3 \approx 2.0$ . Hence, according to determinant *Attitude*, *car* would have the highest chance to be picked for this agent, followed by *train* and *walking*.

### 3.3 An Overview

An agent's main components and their functions are illustrated in Figure 5. When a task is assigned, the *Perception* observes the current state of the environment, other's opinions and combines them with the agent's internal state to produce a list of available options. They are given to the *Decision-making* unit to be evaluated using the functions (or preferences) from the *Memory*. Details of this process are described in Section 3.2. The *Communication* component then utilises this result to create a behaviour output and communicate this with the environment and others. A social agent can also have different objectives to achieve (e.g. influence others' decisions and/or promote a certain action), which might affect the outcome action and their opinions. The environment can then




 Fig. 4: An utility example in mobility domain <sup>1</sup>

provide feedback(s) based on the numbers of demands and nature of the system associated with the action. Agent remembers these past results in the *Memory*, which can then be used to modify the probability of expected values in future decision-making as described in Section 3.1.

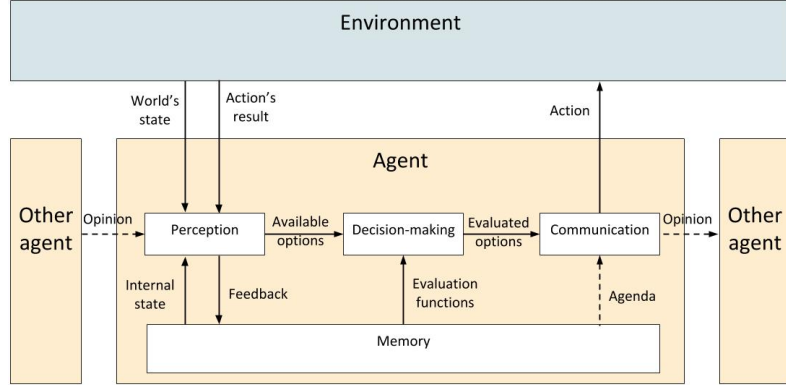


Fig. 5: Overview of agent's architecture

## 4 Experiment

This section focus on the usage of a platform called BedDeM to perform some experimentation regarding the effects of trust and reputation in mobility modal choices. It is being developed in Java based on the Repast library with the main

purpose to generate yearly demands at the individual household level for the population of Switzerland. It is worth noticing that BedDeM is not a routing model and hence, agents only take feedback from the environment as the indication on whether a trip is either a success or a failure. More details of its main functions are described in the next subsection, which is followed by the current empirical mapping. We then consider some testing scenarios that highlight the effects of trust and reputation in mobility modal choices. The final subsection provides some discussion on the competency of BedDeM as a trust and reputation model and suggests some extensions for future development.

#### 4.1 The Behaviour-driven Demand Model (BedDeM)

As shown in Figure 6, BedDeM consists of two processes. In the configuration phase, we applied cluster analysis on the Swiss Household Energy Demand Survey (SHEDS) [36] and the Mobility and Transport Micro-census (MTMC) [2] to generate the mobility profiles. This process is detailed in [4]. A number of random respondents from the surveys that matched these clusters are selected to generate a synthetic agent population for simulation.

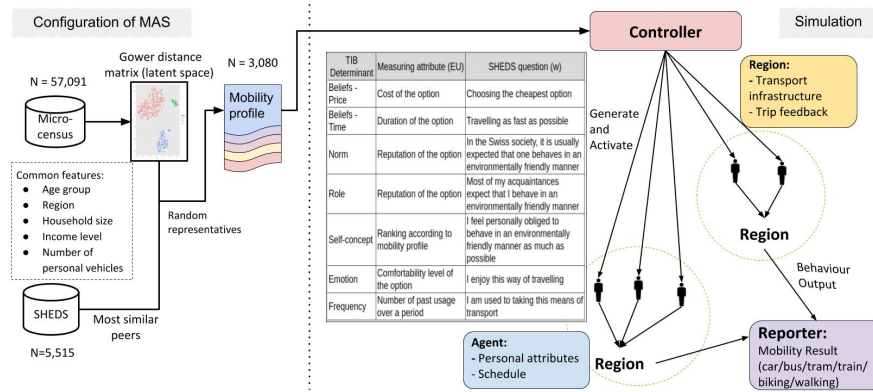


Fig. 6: Agent decision-making mechanism with TIB Module

These data are then passed to a controller in the simulation stage. Each agent contains a mapping of qualitative data in SHEDS to all TIB's determinants, which is calibrated to reproduce the traveling patterns for in MTMC of 2015. A sample of the first level determinants is included in Figure 6. More details of the data mapping and agent's parameters calibration can be found in [25]. An agent represents a typical household in Switzerland and is assigned a weight-to-universe number so that the end results can be scaled back to the national level. Agents from the same cluster are considered neighbors when the reputation of a service needs to be calculated. Depend on the availability of transportation systems, the set of options in decision-making often consists of rail, bus, tram,

car, walking, biking. The agents are then assigned to their respective regions, which contain information about the available public transportation and can be used to reflect the dynamic change in mobility infrastructure (e.g. traffic jams, available public transport) with regard to the simulating time. Each region also has a mechanism to provide feedback after receiving an action as the result of the agent’s decision-making (see Section 4.2).

## 4.2 Setup

The purpose of this experiment is to investigate the effect of a public transportation’s reputation to its demand, which is measured in yearly total kilometers travelled. The reputation in this case represented by a constant percentage, (ground truth value), i.e. punctuality. When the agent performs an action, a different percentage is randomly generated. If its value under the ground truth, the region will output a successful signal to the agent and vice versa. These consequences are connected with the duration of a trip for the agent. Successful feedback means the agent has performed the trip within its time estimation. In contrast, a failure message from the region means running time should be double for this trip. These information are taken into account in the next agent’s decision-making as explained in Section 3.1 and Section 3.2. After decisions of all agents have been made, the region computes the new reputation of the service by joining its residents’ opinions using consensus rule (see Equation 3), while ignoring all empty opinions. Currently, the agents do not have the ability to lie about their opinions and thus, the reputation reflects collective perception about the punctuality of a service. The national level reputation of a service is simply the combination of all regional reputation by also utilizing the consensus rule.

In this study, we focus on changing the successful rate of the rail service operated mainly by Swiss Federal Railways . On one hand, punctuality is an important determinant of quality of service [13] and in turn, affect the yearly demand. Conversely, other determinants (e.g. speed of trains, average fare per km and safety) can also contribute to the customer’s perception about the performance index [13]. Hence, the rail service provides an excellent testing ground to observe the contribution and effects of trust, reputation along with other socio-psychological determinants in the agent’s decision-making process.

A summary of the setup for the experiment can also be found in Table 1. We can divide them into five main categories, whose details are as follow:

- *Perfect world scenario*: The successful rate of all trips is kept at 100%. This setup is similar to what has been done previously in [25]. Therefore, we expect that there’s almost no uncertainty in the agent trust. This can act as a base to compare the effects of trust and reputation when they are implemented in later settings.
- *Real world scenario*: According to [27], around 87.8% of total trips were punctual in 2015, which is the year we calibrated our agents in [25]. We assign this number to all regional rail trips’ successfully rate.

- *Agents starting with low or high expectation scenario*: The initial value of the probability expectation for an option  $x$ ,  $E(\omega_x)$ , will be set at either 0 or 1, instead of  $1/2$  as suggested in Section 3.1. The punctuality of rail service is still kept at 87.8%.
- *Regional disruption scenarios*: The punctuality of rail service at a single region is set at 25%. Otherwise, it is similar to the real-world case, i.e. 87.8%. These structures can be used to test the effect of disruption at one single region to the figures of other regions and the national level.

Table 1: Collective ground truth / Punctuality (measured in percentage) of rail service in individual region for different testing scenarios

Scenario	Region 1	Region 2	...	Region 26
Perfect world	100	100	...	100
Real world	87.8	87.8	...	87.8
High expectation	87.8	87.8	...	87.8
Low expectation	87.8	87.8	...	87.8
Disrupt in Region 1	<b>25</b>	87.8	...	87.8
Disrupt in Region 2	87.8	<b>25</b>	...	87.8
...	...	...	...	...
Disrupt Region 26	87.8	87.8	...	<b>25</b>

### 4.3 Results

As the model contains an element of randomness, we perform 100 simulations for each scenario and compute the averages. Table 2 shows the reputation in the percentage of the rail service at the national level and three representative regions 1, 24 and 18. They are the highest, mid point and lowest in term of kilometres demand in the perfect world scenario. Other disruptive scenarios follow the same pattern as these representatives.

**Reputation:** In essence, reputation measures at the national level and regions that has large train usage (e.g. Region 1) are nearly the same as the ground truth. The lower overall kilometres travelled by agents, the less accurate cumulative reputation of the whole region. This can be seen in the figures of Region 18. When a disruption happens at a single region, the rail’s reputation converse to the ground truth. Despite that, we observe no notable change in the national figures. Except for the disruptive regions, other regional figures also show no significant difference. This indicates the inter-connections between agents from separate regions are weak or limited; and therefore, the effects of interference cannot spread.

Table 2: Result of reputation (measured in percentage) and total demand (measured in kilometres travelled) of rail service at the national level and three representative regions

Scenario	Reputation (percentage)				Total demand (10 <sup>6</sup> km)			
	N	R 1	R 24	R 18	N	R 1	R 24	R 18
Perfect world	<b>99.6</b>	99.9	99.4	95.7	<b>9081.588</b>	3114.568	104.632	0.173
Real world	<b>86.3</b>	89.0	87.8	83.0	<b>8958.592</b>	3112.998	104.632	0.173
High expectation	<b>87.7</b>	87.6	82.5	84.8	<b>16000.392</b>	5002.029	142.432	97.759
Low expectation	<b>85.2</b>	87.3	86.5	74.5	<b>8878.993</b>	3105.089	104.632	0.173
Disrupt Region 1	85.2	<b>25.7</b>	88.2	86.5	8865.848	<b>3105.089</b>	98.867	0.173
Disrupt Region 24	87.4	88.5	<b>22.0</b>	95.7	8937.483	3112.998	<b>102.267</b>	0.173
Disrupt Region 18	86.5	86.7	88.3	<b>12.8</b>	8947.010	3112.178	103.876	<b>0.173</b>

**Total rail-kilometers:** By implementing a lower percentage of trust and reputation in the real world setup, we observe a significant decrease in usage at the national level and major regions (e.g. Region 1). Having agents starting with high probability expectation affects the overall kilometres travelled, with the national figures almost doubled and high increases in all regions. Conversely, agents starting with low expectation cause a small drop in both national and Region 1’s total kilometers. The demand at Region 24 and 18 however, do not exhibit any sizable changes in this scenario. As trust and reputation are only a subset of determinants in Triandis’ Theory, these observations signify that agents have taken into account not only the probability expectation values but also other environmental difference.

There is no extensive impact on the national figures in the disruptive cases compare to the real world scenario. The reduction is noticeable when the disruption happens on the regions that have large demand where punctuation is the key issue there (e.g. Zurich, Geneva). In contrast, the interference does not make any substantial changes to the regions that originally have a small consumption number (i.e. Region 18) since the train is the only available option for these particular residents.

#### 4.4 Discussion

The current preliminary results show that the reputation of the rail system in the agents’ opinions can reflect the ground truth in the setup. As seen in the experiment result, the current mobility model is capable of showing the difference in reputation at two separate levels - regional and national. It suggests a throughout investigation at different angles is needed when a change happens (e.g. disruption, new services and policies). Realistic scenarios with more interference at a variety of scales can also be set up similarly. In addition, the experiment indicates that trust and reputation are two of the many determinants in the agent’s

decision-making; and thus, there are differences when comparing the change in total kilometres demands at the regional level in the high/low expectation and disruptive cases. By changing their weights in Utility Function (see Section 3.2), we also have the opportunity to further explore and compare their contribution to the agent’s decision-making from a micro-level perspective.

There are some limitations in the model that needed to be considered in the future development stage. Firstly, it assumes that the punctuality of the rail service is static for the duration of the simulation. Using more up-to-date data from online sources such as [34], the traffic rate can be adapted dynamically based on the time of the simulation. Punctuality of other available options (e.g. car, bus, tram) is also an essential feature in order to reflect the transition in total demand more accurate. Another approach involves utilising subjective logic on groups of determinants in TIB to derive more sets of complex outcome scenarios for simulation purpose.

## 5 Conclusion

Our main contribution in this paper is the development and implementation of a novel concept using subjective logic to represent trust and reputation elements in a decision-making framework that utilises Triandis’ Theory of Interpersonal Behaviour. Its main purpose is to provide a way to incorporate empirical data in the simulation process and make use of the variety in psychological ontologies to produce a more transparent explanation for social phenomena. Subsequently, this has the potential to facilitate the engagement of psychologists, sociologists, economists and the general public with multiagent modelling projects. We demonstrate the framework practicality by performing experiments aiming at replicating the punctuality of rail service and observing the effect of trust and reputation to the number of kilometres travelled. Although there are still some limitations, the initial results show that the model is fundamentally capable of archiving those targets.

This trust and reputation modelling framework can be further extended by allowing the usage of multinomial and hyper opinions in subjective logic [20]. In other words, the agent can consider the probabilities of the different possible outcomes that are formed by a set of independent determinants (e.g. price, time). Another direction would be to utilise fuzzy logic [26] to exam data in surveys and classify the quality of a trip as either success or failure, especially for qualitative determinants (e.g. emotion, self-concept). Using geographical statistics can also allow us to observe how physical proximity, the layout of regions and other environmental elements can hinder or facilitate interaction and interdependence among social groups. Since *Norm* is incorporated in the agent’s decision-making (see Section 3.2), another potential direction is investigating on how to integrate moral preferences in the agent’s reasoning. In addition, we also aim to use the framework to address the issues in trusted Artificial Intelligent and trust modelling as stated in [11], especially emphasis on creating support for interoperability and employ domain-specific ontologies in simulation research.

## References

1. Ajzen, I.: The theory of planned behavior. *Organizational behavior and human decision processes* **50**(2), 179–211 (1991)
2. ARE/BfS: Verkehrsverhalten der Bevölkerung Ergebnisse des Mikrozensus Mobilität und Verkehr 2015. Federal Office for Spatial Development and Swiss Federal Statistical Office (2017)
3. Axelrod, R., Dion, D.: The further evolution of cooperation. *Science* **242**(4884), 1385–1390 (1988)
4. Bektas, A., Schumann, R.: How to optimize gower distance weights for the k-medoids clustering algorithm to obtain mobility profiles of the swiss population (2019), to be published and represented in the Swiss Conference on Data Science (SDS) Conference - June 2019
5. Bhatnagar, R.K., Kanal, L.N.: Handling uncertain information: a review of numeric and non-numeric methods. In: *Machine Intelligence and Pattern Recognition*, vol. 4, pp. 3–26. Elsevier (1986)
6. Birk, A.: Learning to trust. In: *Trust in Cyber-societies*, pp. 133–144. Springer (2001)
7. Braynov, S., Sandholm, T.: Contracting with uncertain level of trust. *Computational Intelligence* **18**(4), 501–514 (2002)
8. Castelfranchi, C., Falcone, R.: Principles of trust for mas: Cognitive anatomy, social importance, and quantification. In: *Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160)*. pp. 72–79. IEEE (1998)
9. Castelfranchi, C., Falcone, R.: Trust is much more than subjective probability: Mental components and sources of trust. In: *Proceedings of the 33rd annual Hawaii international conference on system sciences*. pp. 10–pp. IEEE (2000)
10. Cho, J.H., Chan, K., Adali, S.: A survey on trust modeling. *ACM Computing Surveys (CSUR)* **48**(2), 28 (2015)
11. Cohen, R., Schaekermann, M., Liu, S., Cormier, M.: Trusted ai and the contribution of trust modeling in multiagent systems. In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. pp. 1644–1648. International Foundation for Autonomous Agents and Multiagent Systems (2019)
12. Dugdale, J.: Human behaviour modelling in complex socio-technical systems: an agent based approach. Ph.D. thesis (2013)
13. Duranton, S., Audier, A., Hazan, J., Langhorn, M.P., Gauche, V.: The 2017 european railway performance index. The Boston Consulting Group (2017)
14. Falcone, R., Castelfranchi, C.: Social trust: A cognitive approach. In: *Trust and deception in virtual societies*, pp. 55–90. Springer (2001)
15. Fishbein, M., Ajzen, I.: Belief, attitude, intention and behavior: An introduction to theory and research (1975)
16. Gambetta, D., et al.: Can we trust trust. *Trust: Making and breaking cooperative relations* **13**, 213–237 (2000)
17. Granatyr, J., Botelho, V., Lessing, O.R., Scalabrin, E.E., Barthès, J.P., Enembreck, F.: Trust and reputation models for multiagent systems. *ACM Computing Surveys (CSUR)* **48**(2), 27 (2015)
18. Henkind, S.J., Harrison, M.C.: An analysis of four uncertainty calculi. *IEEE Transactions on Systems, Man, and Cybernetics* **18**(5), 700–714 (1988)
19. Jøsang, A.: Artificial reasoning with subjective logic. In: *Proceedings of the second Australian workshop on commonsense reasoning*. vol. 48, p. 34. Citeseer (1997)
20. Jøsang, A.: Subjective logic. Springer (2016)

21. Kambhampati, S.: Synthesizing explainable behavior for human-ai collaboration. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1–2. International Foundation for Autonomous Agents and Multiagent Systems (2019)
22. Mui, L., Mohtashemi, M., Halberstadt, A.: A computational model of trust and reputation. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. pp. 2431–2439. IEEE (2002)
23. Mukherjee, R., Banerjee, B., Sen, S.: Learning mutual trust. In: Trust in Cyber-societies, pp. 145–158. Springer (2001)
24. Nguyen, K., Schumann, R.: An exploratory comparison of behavioural determinants in mobility modal choices (2019)
25. Nguyen, K., Schumann, R.: On developing a more comprehensive decision-making architecture for empirical social research: Lesson from agent-based simulation of mobility demands in Switzerland (2019)
26. Novák, V., Perfilieva, I., Mockor, J.: Mathematical principles of fuzzy logic, vol. 517. Springer Science & Business Media (2012)
27. Railways, S.F.: Sbb: Facts and figures 2018. In: Technical Report (2018)
28. Ramchurn, S.D., Huynh, D., Jennings, N.R.: Trust in multi-agent systems. The Knowledge Engineering Review **19**(1), 1–25 (2004)
29. Sabater, J., Sierra, C.: Regret: reputation in gregarious societies. In: Agents. vol. 1, pp. 194–195 (2001)
30. Schillo, M., Funk, P., Rovatsos, M.: Using trust for detecting deceitful agents in artificial societies. Applied Artificial Intelligence **14**(8), 825–848 (2000)
31. Sen, S., Dutta, P.S.: The evolution and stability of cooperative traits. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 3. pp. 1114–1120. ACM (2002)
32. Sen, S., Sajja, N.: Robustness of reputation-based trust: Boolean case. In: Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1. pp. 288–293. ACM (2002)
33. Song, S., Hwang, K., Zhou, R., Kwok, Y.K.: Trusted p2p transactions with fuzzy reputation aggregation. IEEE Internet computing **9**(6), 24–34 (2005)
34. tool, O.: Analyzes on public transport in switzerland. <http://puenktlichkeit.ch> (2019), [Last accessed 16-October-2019]
35. Triandis, H.C.: Interpersonal behavior. Brooks/Cole Pub. Co. (1977)
36. Weber, S., Burger, P., Farsi, M., Martinez-Cruz, A.L., Puntiroli, M., Schubert, I., Volland, B.: Swiss household energy demand survey (sheds): Objectives, design, and implementation. Tech. rep., IRENE Working Paper (2017)
37. Witkowski, M., Artikis, A., Pitt, J.: Experiments in building experiential trust in a society of objective-trust based agents. In: Trust in Cyber-societies, pp. 111–132. Springer (2001)
38. Wu, D., Sun, Y.: The emergence of trust in multi-agent bidding: a computational approach. In: Proceedings of the 34th Annual Hawaii International Conference on System Sciences. pp. 8–pp. IEEE (2001)
39. Yager, R., Fedrizzi, M., Kacprzyk, J.: Advances in the dempster-shafer theory of evidence (1994)
40. Yu, B., Singh, M.P.: Distributed reputation management for electronic commerce. Computational Intelligence **18**(4), 535–549 (2002)
41. Yu, B., Singh, M.P.: An evidential model of distributed reputation management. In: Proceedings of the first international joint conference on Autonomous Agents and Multiagent Systems: Part 1. pp. 294–301. ACM (2002)