ABSTRACT

SUNDARESAN, DIVYA. User Modeling for Prosocial Nudging with Reinforcement Learning. (Under the direction of Dr. Munindar P. Singh).

Prosociality refers to an attitude or behavior that is intended to benefit others. This thesis demonstrates a computational approach to prosociality in the context of a *(public) microtransit* service for disadvantaged riders. Practical resource allocation problems, such as in microtransit, are often over-constrained in that user requests (pickup and dropoff locations at specified times) cannot all be satisfied or satisfied only by violating soft goals such as sustainability.

We posit that instead of taking user preferences as fixed, shaping them toward prosociality will lead to improved societal outcomes. Prosociality appears as a willingness to adjusting one's pickup and dropoff times and locations to accommodate the schedules of others and to enable sharing rides (which increases the riders served with the same resources). Importantly, economic signals are unacceptable in public microtransit. For example, surge pricing would be unethical because the targeted riders lack funds to pay a premium and would be thus coerced into undesirable options.

We use the ACT-R cognitive architecture (Adaptive Control of Thought- Rational) to represent the human mind and model human decision making, and reinforcement learning to learn optimal *nudges* for users in two dimensions: spatial adjustment preferences and empathetic tendencies. We demonstrate our idea with the help of simulations considering a diverse set of users, and an illustration via a mobile app using ArcGIS. We find that using reinforcement learning to understand user preferences helps guide them towards prosociality.

User Modeling for Prosocial Nudging with Reinforcement Learning

by
Divya Sundaresan

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Master of Science

Computer Science

Raleigh, North Carolina
2023

APPROVED BY:

Dr. Eleni Bardaka	Dr. Noboru Matsuda

Dr. Munindar P. Singh
Chair of Advisory Committee

BIOGRAPHY

Divya Sundaresan was born in New Jersey, and lived there until the age of ten. She then moved to Chennai, India, where she stayed until she finished school, after which she received her BTech in Computer Science at National Institute of Technology, Delhi. She worked for a few years at Deloitte before moving back to the US for her MS in Computer Science at NC State University. Her interests currently lie in AI, specifically the development of safe and responsible AI systems. In addition to research, she enjoys dancing and writing.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Munindar P. Singh, for introducing me to the field of research. His encouragement and support inspired me to embark upon a path I would have never considered otherwise. I would like to thank Aditya Khadse for implementing the prototype app and for always being ready to help debug it. I would like to thank my committee members, Dr. Eleni Bardaka and Dr. Noboru Matsuda, for their valuable inputs and feedback. I would also like to thank my family for always supporting, encouraging, and believing in me. Finally, I would like to thank the NSF for partially supporting this research under grant CNS-2125447.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

Microtransit services are a shared, technology-enabled, public transit system with flexible routing and pick-up and dropoff locations developed based on real-time trip demand and origin-destination patterns [Shaheen et al., 2020]. In suburban and rural areas of the US, such services are inefficient and infrequent, if they exist at all [Zimmerman et al., 2015]. In this study, we use AI techniques to determine the best interventions to persuade microtransit users to behave prosocially [Bardaka et al., 2020]. Previous research on nudging and persuasion for transportation and urban mobility [Anagnostopoulou et al., 2018; Kormos et al., 2021] provides valuable insights from behavioral science and ideas on effective nudging techniques, including goal setting, personalized messaging, social comparison, and gamification of the system.

We consider a multistakeholder cyberphysical (sociotechnical) system, or STS Singh [2014] in microtransit services. The stakeholders (including users and providers, i.e., riders, drivers, and the city transit authority) form the social tier and the cyberphysical resources and data (i.e., vehicles and the associated information technology to request rides) form the technical tier. We posit that problems that may be difficult to solve at the technical tier can be made tractable through interventions in the social tier.

We aim to nudge users toward prosocial behavior by suggesting adjustments to their ride requests. To build an agent that suggests acceptable adjustments, our first challenge is to understand our users. Suggestions must be tailored to individual users for maximum effect. Some people are willing to adjust their times but not their pickup and dropoff locations (temporal vs spatial flexibility). For other users, it may be the opposite case. Some people may be willing to walk further at certain times of the day than at others. Similarly, people may have different preferences in terms of who they are willing to adjust their schedules for. All these factors must be considered while suggesting alternatives to produce a positive experience both for the user on the receiving end and for the agent (to increase the likelihood of acceptance by the rider). For example, on a sunny day, a certain user may be willing to walk further for pickup than on a rainy day. Our agent should recognize that a prompt to walk further on a rainy day may only antagonize the user and is unlikely to be accepted. However, the same user may be willing to adjust their time, and this is what we aim to learn and suggest. Similarly, a female user may have a low tolerance for walking on the street at night due to safety concerns. Our agent should not suggest that she compromise for someone else and walk in such a situation, because that is undermining her well being. If a user needs to go somewhere urgently, a suggestion to postpone their ride will not be accepted. One person may be more willing to compromise if doing so would benefit a senior citizen, while another may empathize with neurodiverse individuals. If data about fellow riders could be captured and shared in such a way that privacy constraints are adhered to, this information could be used to persuade users to compromise for someone they perceive to be less fortunate or disadvantaged.

For the purpose of this study, we attempt to learn optimal and persuasive suggestions for microtransit users regarding their spatial adjustment and social preferences. We consider a simplified environment and model users to have different spatial thresholds for different environmental conditions. After considering two major cognitive architectures, ACT-R [Anderson, 1996] and Soar [Laird and Rosenbloom, 1994; Laird, 2022], we decided to adopt ACT-R for this work for practical reasons: the existence of an easy to use Python library [Brasoveanu and Dotlačil, 2020]. We use ACT-R to mimic the operations that enable the human mind and hence simulate realistic human decision-making. We also simulate a user's responsiveness to certain persuasive strategies via *value phrases*. A user's value phrases indicate the communities or causes the user empathizes with. We consider different users to have different orientations within the Social Value Orientation (SVO) ring [Balliet et al., 2009], which affects the intrinsic reward they receive from performing different actions. Then we propose that learning these preferences and values results in the ability to produce an optimized solution to a problem—one that benefits the overall

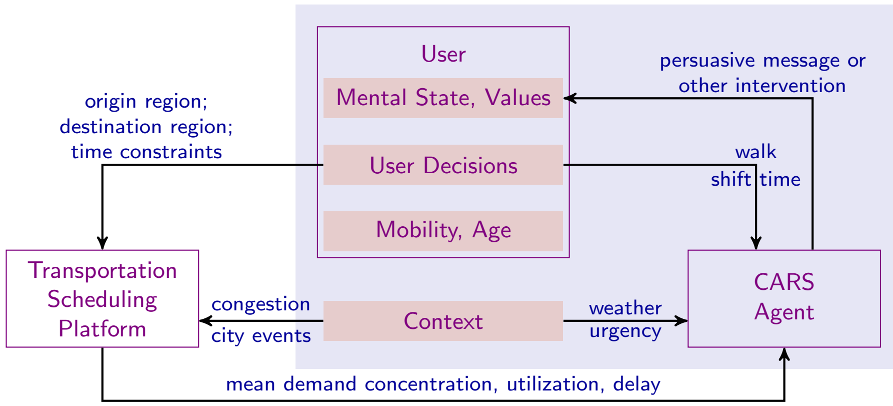


Figure 1.1: Proposed system operation

system and satisfies riders innately as well.

Figure 1.1 shows the proposed operation of the entire system. This work focuses on the CARS (Cooperative Adaptive Ride Sharing) agent which aims to produce the most optimal and persuasive suggestions for a user, given the current conditions and the agent's knowledge of user preferences.

1.1 Research Objective

Our research objective is to understand user decision-making and whether it is possible to nudge humans toward prosocial behavior. The crux of nudging lies in learning user preferences. Accordingly, we identify the following research questions:

RQ1: Can we produce a realistic cognitive model to simulate user decision-making?

RQ2: Can we learn user preferences in terms of how much and for whom or what they are willing to compromise?

RQ3: Can such a model result in a more prosocial ecosystem?

1.2 Contributions

We make three main contributions. Firstly, we describe a reinforcement learning approach for user modeling to determine the most effective interventions to persuade users to behave prosocially. Secondly, we use a cognitive architecture (ACT-R) agent to create a realistic model of human decision-making, as a surrogate for a microtransit user. Finally, we use ArcGIS [Booth et al., 2001], a suite of online geographical information system software, to build a mobile app prototype for microtransit.

1.3 User Studies

Our research goals were co-developed with the city of Wilson, NC which was the first in North Carolina to implement a city-wide microtransit system. Microtransit demand has been rising in small, disadvantaged communities due to its affordability, as microtransit rides are meant to be shared and only charge a nominal fee. Most of the microtransit users in Wilson are low-income and minorities. Wilson's microtransit system facilitates most of the transportation in the community.

Our research team traveled to Wilson in November 2022 to hold a workshop with key stakeholders and revise the research vision, as well as conduct a mock focus group with some of the drivers of Wilson's system (called RIDE). During this visit, we learned about the main uses and deficiencies of the current application. In spite of the extremely high demand experienced by the app, 25% of rides are not completed: out of about 25,000 requests per month, around 18,000 are completed. Currently, RIDE does not support booking in advance, which our team suggested may improve canceling rates, as many users are likely to cancel because of too long a waiting time. Such users would prefer to book in advance for a trip that they cannot be flexible with (such as going to work or a healthcare appointment) as well as a trip they can be flexible with (such as going to a grocery store). Most of the trips scheduled originate from the poorer neighborhoods of Wilson and go to the downtown area.

We also spoke with two RIDE drivers about their experiences with riders and the possibilities of nudging riders. They shared that some riders would be willing to walk, but not, for example, someone with bags of heavy groceries. They also mentioned that pickup and dropoff locations are not always convenient for riders, sometimes creating unsafe crossings and walk requests, as the current application does not consider local traffic and network conditions.

In further focus groups conducted in Wilson by the team, many participants expressed flexibility in their travel schedules and showed compassion for others who they perceived to be more vulnerable, stating that such riders should be prioritized over themselves. They had different degrees of spatial flexibility and different priorities, such as safety or weather. Although many users acknowledged that a rewards program, where people get recognized for their adjustments (such as karma points), could be helpful in persuading people, others claimed that it would not be necessary, commenting that "kindness does not cost anything".

1.4 Organization

The rest of this thesis is organized as follows. Chapter 2 explains a reinforcement learning approach to user modeling, including how we formulate the problem and use user-specified input to model users and learn the most optimal and persuasive interventions for them. Chapter 3 introduces the cognitive architecture ACT-R and explains how we use its python implementation, pyactr, to model human decision-making for the task of accepting or rejecting nudges. Chapter 4 describes experiments with a diverse set of individuals who have different characteristics and preferences, and shows how our approach fares in learning acceptable interventions for them. Chapter 5 demonstrates our idea of prosocial interventions in microtransit with the help of a prototype app. Chapter 6 summarizes our findings and lays out some possible areas of future work in the field.

CHAPTER

2

USER MODELING WITH REINFORCEMENT LEARNING

User preferences play a large role in the suggestions they accept. We aim to learn these preferences in two dimensions: a user's spatial (walking) preferences under different environmental conditions, and the persuasive factors, or *value phrases*, that they respond to. For the purposes of this study, we consider a simplified environment with two features:

1. Weather: sunny or rainy

2. Time of day: morning, afternoon or evening

We presume that users have certain spatial thresholds for each of the six environmental combinations produced from these features. We also assume users to have specific persuasive value phrases, corresponding to the categories of people or the general factors that the user empathizes with. We describe a reinforcement learning approach for user modeling that helps us learn user preferences in these two dimensions.

2.1 Spatial Adjustment Learning

We use model-free reinforcement learning (Proximal Policy Optimization) to learn optimal spatial suggestions for users. We experiment with traditional as well as modified model-free reinforcement learning. For modified model-free reinforcement learning, we use a base model which we pretrain on user profile data. This provides a non-naive starting model that the user can interact with. Our traditional model-free reinforcement learning agent is trained directly through interactions with the user, i.e., without pretraining it on user profile data. Our model aims to learn how far a user would be willing to walk under different circumstances. We reward the model based on the magnitude of accepted spatial adjustment so that it learns to suggest the largest possible adjustment that is likely to be accepted.

2.1.1 Base Model

There are some preferences that users can self-specify. For example, most people may be willing to walk further in the sun than in the rain. However, humans are notoriously incapable of *quantifying* their preferences. We suggest an alternative mode of initial input on the part of the user: a feature trace [Bobu et al., 2022] specifying their weather preferences (for example, *sunny* > *rainy*), and their time of day preferences (for example, *morning, afternoon* > *evening*). We also consider other data provided by the user (their age and gender). Using this user profile, we create a base ordering of environmental conditions and define default spatial thresholds for them. We train a base model on these default values. An example of computed base thresholds based on feature traces is as follows:

Listing 1: Feature traces

```
Weather feature trace: [sunny > rainy]
Time of Day feature trace: [afternoon > morning > evening]
Gender: female
Age: >65
```

The resultant initial thresholds set would be:

Listing 2: Thresholds computed based on feature traces

```
Sunny afternoon: 1 (sunny) * 450 (default) * 1 (gender) * 0.75
    (age)
```

```
Sunny morning: 1 (sunny) * 300 (default) * 1 (gender) * 0.75 (age)
Sunny evening: 1 (sunny) * 100 (default) * 0.5 (gender) * 0.75 (age)
Rainy afternoon: 0.5 (rainy) * 450 (default) * 1 (gender) * 0.75 (age)
Rainy morning: 0.5 (rainy) * 300 (default) * 1 (gender) * 0.75 (age)
Rainy evening: 0.5 (rainy) * 100 (default) * 0.5 (gender) * 0.75 (age)
```

2.1.2 Main Model

Once we have trained the base model on the computed base thresholds for a certain user (for example, in the above case, a $>$ 65 female with the self-specified feature traces mentioned above), we use that model as the starting point which the user actually interacts with. This way, our model starts out with some knowledge and is not completely naive. For example, it knows not to suggest higher distances to walk when it is raining, as it has been trained with that information. Over time, the model fine-tunes itself and learns actual user thresholds, enabling it to make spatial adjustment suggestions that the user is likely to accept. We also train a traditional model-free reinforcement learning agent directly, i.e., without pretraining it on user profile data. This model has no initial knowledge, but learns through interactions with the user. We use Proximal Policy Optimization (PPO) [Schulman et al., 2017] to train both the base and main model.

2.2 Persuasion Learning

The second preference we wish to learn is *who* and *what* people are willing to adjust for. We define a set of *value phrases* that a user may be empathetic toward. We consider a set of people: *babies*, *children*, *senior citizens*, *ill people*, *people with disabilities*, and *neurodiverse individuals*. In addition, we consider general factors: *environmental benefit* and *health benefit*. Each user may resonate with a subset of these value phrases.

During every simulated ride, we simulate current conditions of the environment, i.e., the fellow riders. Given the fellow riders, we want to learn which of them to mention to our user to maximize the chance of them accepting our suggestion. If the user rejects the

Table 2.1: ACT-R Hyperparameters

Parameter	Value
subsymbolic	True
utility noise	0.01
utility learning	True
utility alpha	0.01
strict harvesting	True

Table 2.2: PPO Hyperparameters

Parameter	Value
policy	MlpPolicy
learning rate	0.0003
number of steps	2048
batch size	64
number of epochs	10
verbose	1

spatial adjustment, we attempt to persuade them further by suggesting who or what they would be helping by making this compromise. We use a multi-armed bandit algorithm to learn the user's empathetic preferences based on their response to this second prompt. For example, *Would you be willing to walk x meters further to location A? You would be helping a senior citizen with their trip.* Or, *By walking x meters to location B, you would be getting y steps, the first step to healthy living!* As stated above, different users may sympathize with different individuals and causes. Over time we learn these value phrases, and so we know what to say to the user to maximize our chances of them accepting the suggestion.

2.3 Hyperparameters

While running our simulation, we use the hyperparameters specified in Tables 2.1, 2.2 and 2.3.

Table 2.3: Multi-Armed Bandit Hyperparameters

Parameter	Value
exploration rate	0.3
uniform distribution parameter	0.2
decay rate	0.8
reward weight	0.8
initial weight	0.5

CHAPTER

3

COGNITIVE MODELING

To build a realistic model of human decision-making, we use a cognitive architecture. A cognitive architecture is a psychological theory aimed to define human cognition at a generic level [Anderson and Lebiere, 2003]. It acts as a framework upon which specific tasks can be defined. The underlying psychological components are shared between tasks. A cognitive architecture explains the processes taking place in the mind during the complete execution of a task: from perception to response. Hence, a cognitive architecture can be thought of as a unified theory of the mind [Newell, 1994]. The two most successful cognitive architectures are ACT-R [Anderson, 1996; Anderson et al., 2004] and Soar [Laird, 2019; Laird and Rosenbloom, 1994]. Laird [2022] provides a comparison between ACT-R and Soar. Due to the existence of an easy-to-use Python library for ACT-R [Brasoveanu and Dotlačil, 2020], we adopt ACT-R for this study. We define a model for decision-making regarding spatial adjustments in microtransit using ACT-R.

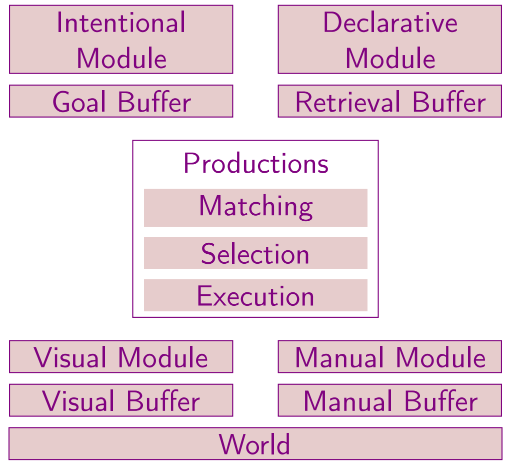


Figure 3.1: Modular software architecture of the cognitive architecture ACT-R.

3.1 Overview of ACT-R

ACT-R (Adaptive Control of Thought—Rational) [Anderson, 1996] is a cognitive architecture which aims to model human cognitive processes in tasks such as perception, attention, memory, reasoning, and problem solving. In ACT-R, the human mind is modeled as a set of three components or cognitive modules, which work together to process information and produce behavior. These three components are the *declarative memory*, which stores facts; the *procedural memory*, which defines actions; and the *control system*, which coordinates the interaction between the declarative and procedural memory and generates behavior according to environmental state. Figure 3.1 describes ACT-R schematically. Data in ACT-R is represented in the form of chunks, mimicking the way humans process, store and retrieve small pieces of information.

ACT-R has been used to model a wide range of cognitive phenomena, including language processing [Ball et al., 2010; Ball, 2013], problem-solving [Lebiere et al., 1998; Peebles and Cheng, 2002], and decision-making [Martin et al., 2004; Kennedy and Patterson, 2012; Dimov et al., 2020]. It has also been applied in various domains, including education [Ritter et al., 2007; Anderson and Schunn, 2013], human-computer interaction [Byrne, 2001; Trafton et al., 2013], and cognitive psychology [Smart and Sycara, 2014; Smart et al., 2014; Borst and Anderson, 2015].

3.2 Decision-Making Using ACT-R

Human decision-making in ACT-R works by way of utility values assigned to different productions. The procedural memory defines the available productions and their prerequisites and consequent states. Given the world state (represented by certain chunks in the goal buffer), we define productions (actions) that can be carried out.

ACT-R chooses the production based on utilities—the one that is likely to provide maximum reward to the agent. It does not always choose the production with the highest utility, aiming to mimic human inconsistency and irrationality, as well as maintain an exploration–exploitation balance. After a production is fired, ACT-R carries out the transitions until the final state, and the utilities of the productions leading up to that state are updated based on the final reward. Over time, the productions which lead to a reward have a higher utility, and the model prefers those productions given a choice of multiple productions for a certain set of chunks in the goal buffer. Preuß [2018] provides a useful overview on learning using ACT-R in the case of Sudoku.

Utility updates in ACT-R occur according to the following formula:

$$U(i, t+1) = U(i, t) + \alpha[R(t) - U(i, t)]$$

where:

$U(i, t)$ is the current utility value of production rule i at time t.

$U(i, t+1)$ is the updated utility value of production rule i at time $t+1$.

α is the learning rate, which determines the extent to which the utility value is updated in response to feedback from the environment.

$R(t)$ is the reinforcement signal or reward received by the system at time t, which reflects the success or failure of the action taken based on the production rule.

$[R(t) - U(i, t)]$ is the prediction error, which represents the difference between the expected reward and the actual reward received by the system.

3.3 Social Value Orientation

Social value orientation (SVO) is a concept in social psychology which states that different individuals have different preferences regarding the allocation of resources between themselves and others [Lange, 1999]. Studies such as those conducted by Balliet et al. [2009] show that social value orientation has a significant effect on behavior in social dilemmas.

We consider archetypal social value orientations, shown as points on the circumference of the circle in Figure 3.2, while modeling users. A person's orientation in the SVO ring is reflected in the internal value they acquire from performing actions. For example, someone with a high self interest and competitiveness would prioritize their outcomes and prefer others' outcomes to be worse than theirs. Someone with a high other interest would care more about the outcomes of others. On the other hand, someone with a high self interest and a high prosociality would prioritize their outcomes, but would prefer that others also get what they want. Correspondingly, the internal value a user gets from an action changes based on these parameters. There are multiple ways to model a user's social value orientation, such as angle in the SVO circle, Sawyer's Altruism Scale, the 9-Item Triple Dominance Scale and Rank Correlation with Decomposed Games [Murphy and Ackermann, 2014].

We model a user's SVO using two parameters.

Other-interest is the degree to which the user values the outcome of others relative to their own. It ranges from 0 (no interest in the outcome of others) to 1 (only interested in the outcome of others). The complement of other-interest is *self-interest* (where self-interest=1-other-interest)

Prosociality is the degree to which the user values the sum of outcomes for themselves and others. It ranges from 0 (completely competitive) to 1 (completely prosocial). The complement of prosociality is *competitiveness* (where competitiveness=1-prosociality)

3.4 Persuasive Value Phrases

We presume that users have preconceived notions regarding whom they find disadvantaged or less fortunate. For the purposes of this study, we consider a few categories of people: *babies*, *children*, *senior citizens*, *ill people*, *people with disabilities*, and *neurodiverse individuals.*

In addition, we consider two general factors: *environmental benefit* and *health benefit* (due to saving fuel costs and walking, respectively) which always apply. We assume that each user may be empathetic toward a subset of these value phrases.

To preserve realism, we encode users to have responses based on similarity to the causes they believe in. For example, if they sympathize with children, they would be

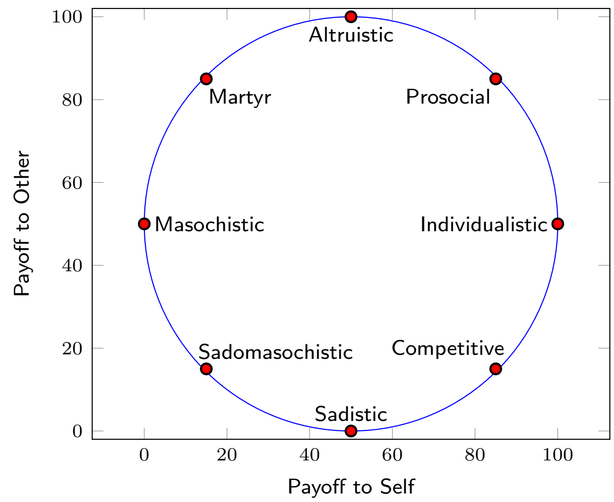


Figure 3.2: A graphical representation of the SVO framework [Murphy and Ackermann, 2014]

more likely to be empathetic towards babies as well. For this, we use a language model to calculate the cosine similarity between the value phrase suggested and the value phrases the user responds to the most.

3.5 Spatial Thresholds

For the purposes of this study, we model a simplified environment based on the following features:

1. Weather: sunny or rainy
2. Time of day: morning, afternoon or evening

We model users to have certain spatial thresholds for each of these environmental combinations. We consider a simple input from each user, allowing them to define their weather and time of day preferences as an ordering, as described in Section 2.1.1. Users' spatial thresholds are the intrinsic thresholds that users have for different environmental conditions. Internal rewards users acquire from different productions depend on these thresholds. A person with a threshold of 100 meters in a certain environmental state would respond differently to an adjustment of 120 meters compared to someone with a threshold of 50 meters, since the amounts of adjustment beyond their thresholds they would have to make are different. We use this assumption to define rewards for our productions, combined with other factors.

3.6 Production Reward Definition

For a realistic utility update of productions, we must define rewards that accurately reflect the internal value a person assigns to a certain production firing, i.e., the satisfaction they acquire. We model this process as goal-directed choice or value-based decision-making Rangel [2009]. Users make decisions based on the comparison of utilities, reflecting the rewards they are likely to receive from their choices. Valuations of choices are compared and desirability of the outcomes that follow are considered to make a decision. The human brain has the capacity to make good (rational) choices for the most part, but not always. ACT-R models this irrationality as utility noise in its decision-making by not always selecting the action with the highest utility.

For the purposes of this study, we consider three cases of a spatial adjustment suggestion, corresponding to a low, medium, and high amount of adjustment on the part of the user. We assume that the reward achieved by accepting these suggestions is inversely proportional to the amount of adjustment. The reward attained by rejecting suggestions is directly proportional to the amount of adjustment: people are likely to feel worse about rejecting interventions that inconvenience them less. In other words, we assume that:

$$\begin{aligned}
\text{reward[low-adjustment-accept]} &> \text{reward[medium-adjustment-accept]} \\
&> \text{reward[high-adjustment-accept]} \\
\text{reward[low-adjustment-reject]} &< \text{reward[medium-adjustment-reject]} \\
&< \text{reward[high-adjustment-reject]}
\end{aligned}$$

We calculate the reward with the following formula.

$$\text{reward} = \begin{cases} 1/a \times o \times p \times m \times (1-u) & \text{if accept} \\ a \times (1-o) \times (1-p) \times (1-m) \times u & \text{if reject} \end{cases} \tag{3.1}$$

where

1. a indicates the *adjustment*, i.e., the fraction that the user would need to walk extra above their threshold.

2. o indicates the *other-interest*, i.e., the degree to which the user values the outcome of others relative to their own; c ranges from 0 (no interest in the outcome of others) to 1 (only interested in the outcome of others).

3. p indicates the *prosociality* of the user, indicating the degree to which the user values the sum of outcomes for themselves and others; p ranges from 0 (completely competitive) to 1 (completely prosocial).

4. m indicates the *mood* of the user, affecting their response to a certain suggestion at a certain time. This adds an additional factor of randomness in the response, mimicking the unpredictable nature of human behavior. It ranges from 0 (completely negative) to 1 (completely positive).

5. u indicates the *urgency* of the current trip. It ranges from 0 (not urgent, implying more flexibility on behalf of the user) to 1 (extremely urgent).

CHAPTER

4

EXPERIMENTS AND RESULTS

To address our research questions, we run experiments with five diverse individuals. We aim to learn user preferences in terms of both acceptable spatial suggestions and persuasive value phrases. A user's characteristics (their SVO orientation) and their current state while requesting a ride (their mood and the urgency of the ride) affect their response to suggested adjustments in addition to their actual spatial thresholds and persuasive value phrases.

4.1 Rider Personas

We consider four archetypal personalities according to SVO research as well as one moderate persona to offset the extremities. We consider that prosociality tends to increase with age [Matsumoto et al., 2016] while modeling these personalities. We also model users above the age of 65 to have slightly lower spatial thresholds for all times of the day, and females to have lower spatial thresholds in the evening. A summary of our users is shown in Table 4.1.

Table 4.1: Personas

SVO	Gender	Age	Other-interest	Prosociality
Moderate	Female	Mid 40s	0.45	0.5
Competitive	Male	Early 30s	0.01	0.01
Individualistic	Female	16	0.01	0.5
Prosocial	Male	Early 50s	0.5	0.99
Altruistic	Female	Early 70s	0.99	0.99

4.2 Hypotheses

4.2.1 Hypotheses

Hypothesis 1 We can create a realistic model of human decision-making using a cognitive architecture.

Hypothesis 2 Optimal spatial suggestions can be learned over time.

Hypothesis 3 Persuasive value phrase suggestions can be learned over time.

Hypothesis 4 Knowledge about a user's basic profile will lead to faster learning.

4.3 Evaluation Metrics

We aim to learn two types of preferences for users: their acceptable spatial adjustments under different conditions and the people or conditions they empathize with. Our agent combines this knowledge to persuade users to accept spatial adjustments and behave prosocially.

4.3.1 Spatial Adjustment Learning Agent

To evaluate the performance of our spatial adjustment learning agent, we use the *accepted spatial adjustment per episode* as our evaluation metric, where episodes are of a fixed length of 1,000 timesteps. This encourages the agent to suggest as high an adjustment as possible that is likely to be accepted by the user. Hence an increasing accepted spatial

Table 4.2: Competitive User: User Profile

Parameter	Value
gender	male
age	early 30s
weather feature trace	sunny > rainy
time of day feature trace	morning = afternoon = evening

adjustment per episode implies that our agent is learning to suggest higher adjustments that are also accepted by our user.

We compare the performance of the untrained model, pretrained model (trained using user profile data) and trained model on our simulated user in terms of the average accepted spatial adjustment, the average percentage of accepted suggestions per episode, and Cohen's d for measuring the standardized effect of using the trained model over the pretrained and untrained models in terms of suggestion quality.

4.3.2 Persuasive Value Phrase Learning Agent

We train our persuasive value phrase learning agent by suggesting a persuasive message whenever a user does not accept a suggestion. This persuasive message includes a value phrase derived from the conditions of the environment and fellow riders on the user's ride. We simulate fellow rider conditions based on a probability distribution in this study. Over time, our agent should learn to suggest the most persuasive value phrase for the user given the current true condition set. We evaluate the performance of our agent by calculating the final ordering of value phrases it predicts as what the user responds to, and comparing that with the user's actual preferences.

4.4 Competitive User

In our first experiment, we consider a completely competitive male in his early 30s. Competitive people seek to maximize the difference between their outcomes and those of others, i.e., to maximize their outcomes while minimizing others'. We consider the shared values specified in Table 4.2 and the internal values specified in Tables 4.3 and 4.4.

Table 4.3: Competitive User: Spatial Thresholds

Weather	Time of Day	Spatial Threshold (in meters)
sunny	morning	300
sunny	afternoon	300
sunny	evening	300
rainy	morning	50
rainy	afternoon	50
rainy	evening	50

Table 4.4: Competitive User: Persuasive Value Phrases

Value Phrase	Persuasive Percentage
health benefit	100 %

4.5 Individualistic User

In our second experiment, we consider a completely individualistic 16 year old female. Individualistic people are concerned only with maximizing their own outcomes. They are neither competitive nor prosocial, only self-interested. In this case, we consider the shared values specified in Table 4.5 and the internal values specified in Tables 4.6 and 4.7.

Table 4.5: Individualistic User: User Profile

Parameter	Value
gender	female
age	16
weather feature trace	sunny > rainy
time of day feature trace	[morning = afternoon] > evening

Table 4.6: Individualistic User: Spatial Thresholds

Weather	Time of Day	Spatial Threshold (in meters)
sunny	morning	450
sunny	afternoon	450
sunny	evening	20
rainy	morning	150
rainy	afternoon	150
rainy	evening	10

Table 4.7: Individualistic User: Persuasive Value Phrases

Value Phrase	Persuasive Percentage
ill people	40 %
senior citizens	36 %
environmental benefit	24 %

4.6 Prosocial User

In our third experiment, we consider a completely prosocial male in his early 50s. Prosocial people prefer mutually beneficial outcomes, considering both their own outcomes and the outcomes of others. We consider the shared values specified in Table 4.8 and the internal values specified in Tables 4.9 and 4.10.

Table 4.8: Prosocial User: User Profile

Parameter	Value
gender	male
age	early 50s
weather feature trace	sunny > rainy
time of day feature trace	[morning = evening] > afternoon

Table 4.9: Prosocial User: Spatial Thresholds

Weather	Time of Day	Spatial Threshold (in meters)
sunny	morning	475
sunny	afternoon	200
sunny	evening	400
rainy	morning	250
rainy	afternoon	200
rainy	evening	20

Table 4.10: Prosocial User: Persuasive Value Phrases

Value Phrase	Persuasive Percentage
environmental benefit	51 %
people with disabilities	49 %

4.7 Altruistic User

In our fourth experiment, we consider a completely altruistic female in her early 70s. Altruistic people are, in addition to being prosocial, selfless as well—they are willing to sacrifice their own well-being for the benefit of others. We consider the shared values specified in Table 4.11 and the internal values specified in Tables 4.12 and 4.13.

Table 4.11: Altruistic User: User Profile

Parameter	Value
gender	female
age	early 70s
weather feature trace	sunny > rainy
time of day feature trace	morning > afternoon > evening

Table 4.12: Altruistic User: Spatial Thresholds

Weather	Time of Day	Spatial Threshold (in meters)
sunny	morning	350
sunny	afternoon	300
sunny	evening	50
rainy	morning	100
rainy	afternoon	100
rainy	evening	20

Table 4.13: Altruistic User: Persuasive Value Phrases

Value Phrase	Persuasive Percentage
babies	55 %
children	45 %

4.8 Moderate User

In our final experiment, we consider a moderate person, who prioritizes their outcomes slightly more than others', and is neither prosocial nor competitive. We consider the shared values specified in Table 4.14 and the internal values specified in Tables 4.15 and 4.16.

Table 4.14: Moderate User: User Profile

Parameter	Value
gender	female
age	mid 40s
weather feature trace	sunny > rainy
time of day feature trace	morning > afternoon > evening

Table 4.15: Moderate User: Spatial Thresholds

Weather	Time of Day	Spatial Threshold (in meters)
sunny	morning	475
sunny	afternoon	400
sunny	evening	50
rainy	morning	250
rainy	afternoon	200
rainy	evening	20

Table 4.16: Moderate User: Persuasive Value Phrases

Value Phrase	Persuasive Percentage
babies	40 %
health benefit	60 %

4.9 Results

For each experiment, we show results for spatial threshold learning and persuasive value phrase learning.

4.9.1 Spatial Adjustments

We use user profile data to hard-code base thresholds for each user, which are used to pretrain a base model, as described in Section 2.1.1.

We then train both our pretrained base model and an untrained model through interactions with our user. For each experiment, we consider pretrained models trained for 50,000 and 100,000 iterations. We define reward for the spatial adjustment learning agent as *magnitude of accepted adjustment*. It can be seen that for each experiment, all the models converge to the same final value, which depicts the maximum amount of spatial adjustment that this user accepts per episode (1,000 time steps).

For the moderate user, the agent learns over time that they are willing to adjust more than is stated by their user profile. This is because they only value their outcome slightly more than others', and are hence willing to compromise for the benefit of others.

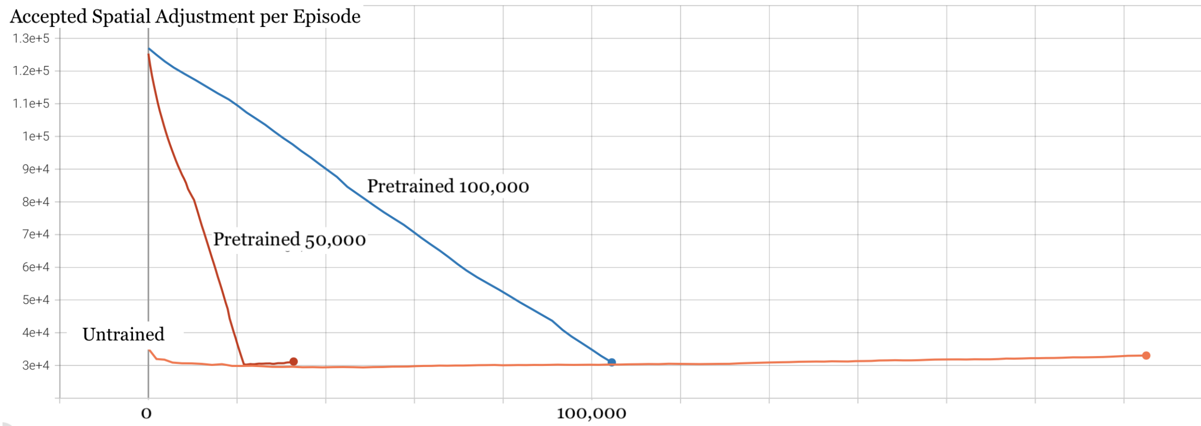


Figure 4.1: Competitive User: Sum of accepted spatial adjustments over fixed length episodes of length 1,000 time steps for pretrained and untrained models.

Hence, the pretrained agents' reward increases over time and all three models converge to the same final value in Figure 4.5. For the competitive user and the individualistic user, the agent's reward remains low because competitive and individualistic people obtain higher internal reward from rejecting suggestions that don't benefit them. An individualistic person obtains satisfaction when their outcomes are better, irrespective of others' outcomes. A competitive person obtains satisfaction when the *difference in outcomes* between themselves and others is maximized, hence they are likely to reject a suggestion even if it does not affect them, since they obtain pleasure from the minimization of the outcomes of others. Correspondingly, the steady state is lower in Figure 4.1 than in Figure 4.2. Our pretrained model does not consider SVO orientation, and so when trained with the competitive or the individualistic personas, the accepted adjustment steadily decreases until it converges to a steady state. The accepted adjustment per episode starting with pretrained models trained for 50,000 and 100,000 iterations on user input data, and an untrained model, are shown as three curves that converge for both the competitive user in Figure 4.1 and the individualistic user in Figure 4.2.

Results for the prosocial and the altruistic user are shown in Figures 4.3 and 4.4. The agent's reward is much higher in these cases because prosocial and altruistic people obtain internal satisfaction from helping other people. However, there is a limit they are willing to compromise, and that is the convergence point for all the models in both cases. Our pretrained model does not consider personality types and therefore the agent slowly learns that these users are willing to compromise much more than is stated in their user profile data. It can be seen that the final agent reward in Figures 4.3 and 4.4 is higher than that

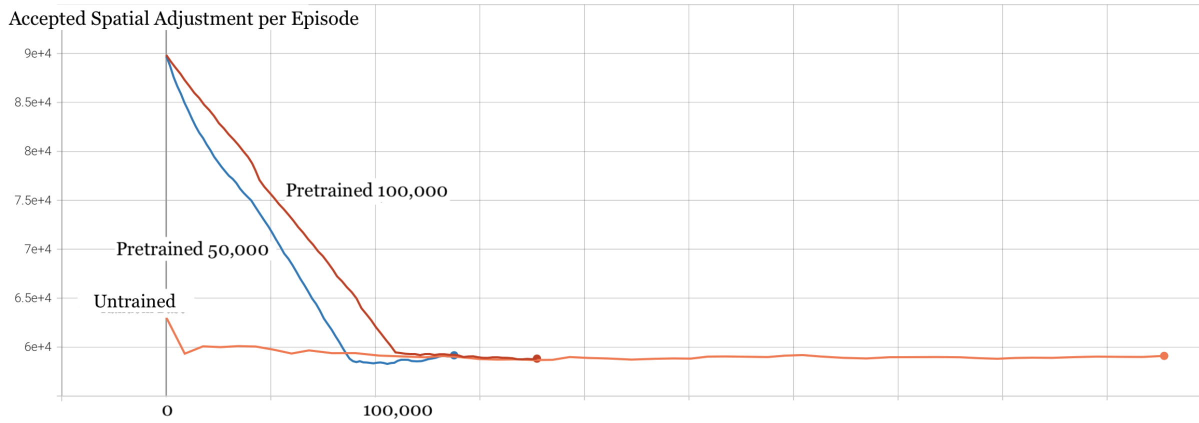


Figure 4.2: Individualistic User: Sum of accepted spatial adjustments over fixed length episodes of length 1,000 time steps for pretrained and untrained models.

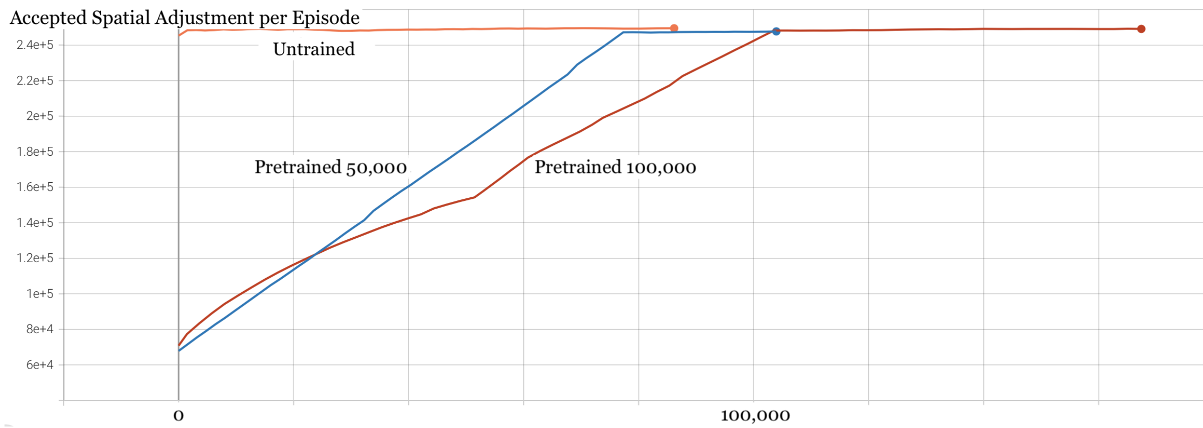


Figure 4.3: Prosocial User: Sum of accepted spatial adjustments over fixed length episodes of length 1,000 time steps for pretrained and untrained models.

of the moderate user in Figure 4.5.

As shown in these figures, our pretrained model takes much longer to converge to the steady state than an untrained model. This is because this model is pretrained with user profile data and does not consider many relevant factors that affect user response. It has to adapt itself to the actual user, who has preferences that are not shared with the agent. Hence, we cannot say that a pretrained agent learns optimal suggestions more quickly than an untrained agent.

However, if we compare the performances of our pretrained model trained for 50,000 iterations with an untrained one in terms of suggestion quality, we can see that our pretrained model performs slightly better. Tables 4.17, 4.18, 4.19, 4.20, and 4.21 provide

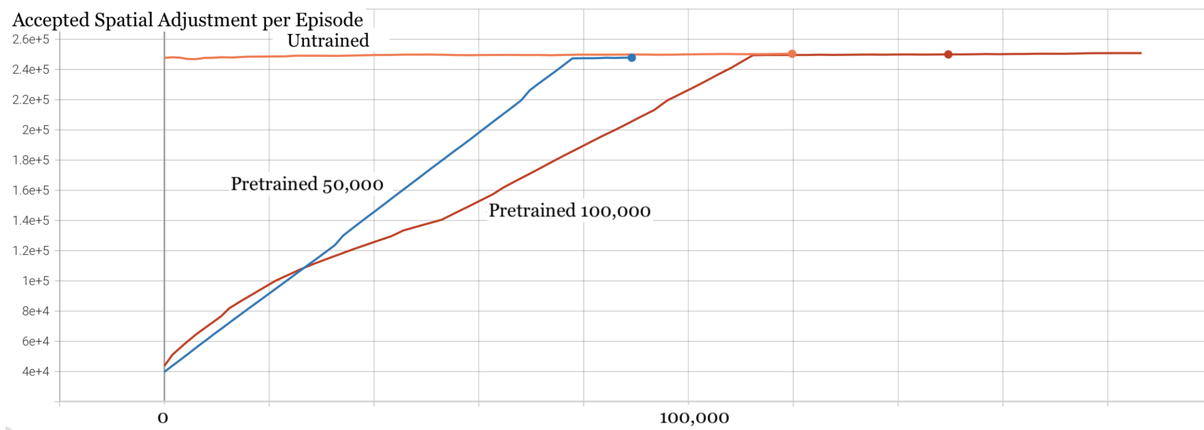


Figure 4.4: Altruistic User: Sum of accepted spatial adjustments over fixed length episodes of length 1,000 time steps for pretrained and untrained models.

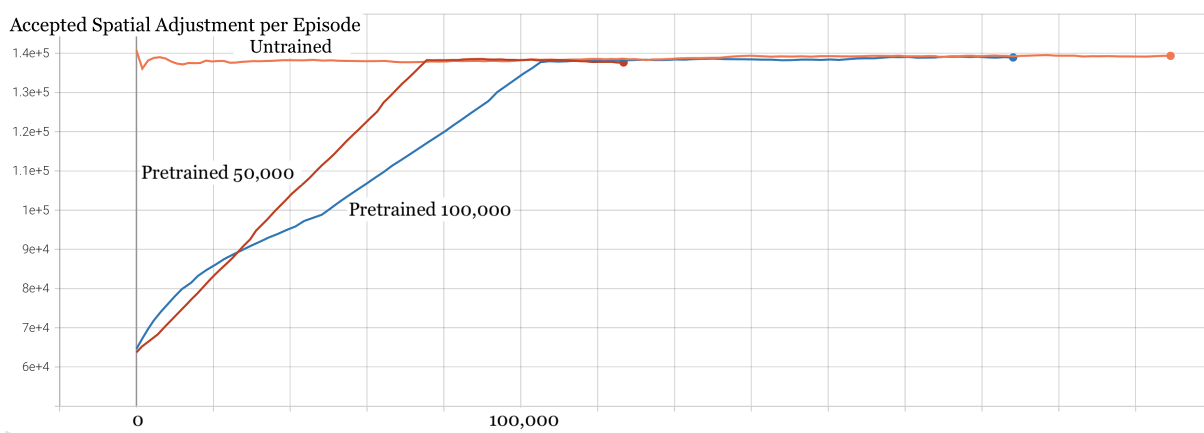


Figure 4.5: Moderate User: Sum of accepted spatial adjustments over fixed length episodes of length 1,000 time steps for pretrained and untrained models.

a comparison between trained, pretrained (for 50,000 iterations) and untrained model performances for each user in terms of suggestion quality. Table 4.22 shows the performance of the trained model with respect to the pretrained and untrained models in terms of the standardized mean difference (Cohen's d) for accepted adjustment per episode and number of acceptances per episode. The trained and pretrained models perform on average better in terms of mean accepted adjustment value per time step and the average percentage of acceptances per episode, than the untrained model. If more data, or more accurate data, could be provided as part of the user profile, our starting model may perform even better. Currently, only a few conditions are considered and it still performs better than an untrained model.

Finally, although the accepted spatial adjustment per episode saturates for all experiments at a certain value, the performance of our trained model is not significantly better than the untrained model. This could be due to the high randomness we model in the ACT-R user, where factors which affect user response, such as their mood and urgency, change with every request and have no way of being known to the agent (unless they are user-specified, but that is unlikely to be accurate). If more information is provided to our agent, especially that which influences user response, the agent is likely to learn a better policy. Future work could explore the effects of less randomness on behalf of the ACT-R agent, or more shared information, on the final performance of the agent.

Another reason for the saturating accepted spatial adjustment could be the reinforcement learning agent's current reward function, which considers the magnitude of the accepted adjustment. Our reasoning behind this reward function is that if it is optimized, the agent should learn the highest possible adjustment it can suggest to a user that the user accepts. However, another reward function might capture the behavior we wish to generate more optimally. Future work could explore different reward functions and compare their performance based on the metrics defined in Section 4.3.1. Future work could also consider different metrics upon which the model can be better evaluated. Although the metrics we define (average accepted adjustment value per episode, average percentage of acceptances per episode) should provide an idea of the performance of the agent, there are flaws in these metrics. For example, a few extremely high suggestions which are unlikely to be accepted could be accepted due to mood, urgency of the trip, or ACT-R's randomness. Similarly, reasonable suggestions could be rejected. Such occurrences affect our metrics, although they are not a reflection of the agent's performance, but rather an indication of the inherent randomness in our users. A better metric might measure model performance more accurately.

Table 4.17: Competitive User: Model evaluation and comparison of average performance over three episodes of length 1,000 time steps.

Metric	Trained	Pretrained	Untrained
Mean Accepted Adjustment Value Per Time Step	73.31	65.49	54.84
Mean Percentage of Acceptances Per Episode	33.00	31.33	30.67
Accepted Adjustment Value Standard Deviation	137.64	130.15	110.20
Percentage of Acceptances Standard Deviation	0.82	2.05	5.73
Percentage Accepted Adjustment Value Increase	33.67	19.42	–
Percentage of Acceptances Increase	7.61	2.17	–

Table 4.18: Individualistic User: Model evaluation and comparison of average performance over three episodes of length 1,000 time steps.

Metric	Trained	Pretrained	Untrained
Mean Accepted Adjustment Value Per Time Step	61.05	54.01	51.49
Mean Percentage of Acceptances Per Episode	31.33	25.33	26.33
Accepted Adjustment Value Standard Deviation	122.35	117.24	107.02
Percentage of Acceptances Standard Deviation	1.24	1.88	3.30
Percentage Accepted Adjustment Value Increase	18.57	4.89	–
Percentage of Acceptances Increase	18.99	−3.80	–

Table 4.19: Prosocial User: Model evaluation and comparison of average performance over three episodes of length 1,000 time steps.

Metric	Trained	Pretrained	Untrained
Mean Accepted Adjustment Value Per Time Step	234.56	211.75	203.04
Mean Percentage of Acceptances Per Episode	93.33	88.00	91.33
Accepted Adjustment Value Standard Deviation	149.01	157.65	148.75
Percentage of Acceptances Standard Deviation	2.87	2.94	6.60
Percentage Accepted Adjustment Value Increase	15.53	4.29	–
Percentage of Acceptances Increase	2.19	−3.65	–

Table 4.20: Altruistic User: Model evaluation and comparison of average performance over three episodes of length 1,000 time steps.

Metric	Trained	Pretrained	Untrained
Mean Accepted Adjustment Value Per Time Step	237.40	226.76	220.21
Mean Percentage of Acceptances Per Episode	92.00	90.00	90.67
Accepted Adjustment Value Standard Deviation	150.50	157.48	157.71
Percentage of Acceptances Standard Deviation	5.10	9.27	10.37
Percentage Accepted Adjustment Value Increase	7.80	2.97	–
Percentage of Acceptances Increase	1.47	−0.74	–

Table 4.21: Moderate User: Model evaluation and comparison of average performance over three episodes of length 1,000 time steps.

Metric	Trained	Pretrained	Untrained
Mean Accepted Adjustment Value Per Time Step	146.58	139.02	130.47
Mean Percentage of Acceptances Per Episode	64.33	64.00	62.67
Accepted Adjustment Value Standard Deviation	156.71	152.03	143.30
Percentage of Acceptances Standard Deviation	1.25	4.32	4.50
Percentage Accepted Adjustment Value Increase	12.35	6.55	–
Percentage of Acceptances Increase	2.66	2.13	–

Table 4.22: Effects of the trained model as Cohen's d scores over pretrained and untrained models for accepted adjustment value and acceptance percentage effect over three episodes of length 1,000 time steps.

Experiment	Metric	Vs Pretrained	Vs Untrained
Competitive	Accepted Adjustment Value	0.03	0.07
Competitive	Percentage of Acceptances	0.58	0.35
Individualistic	Accepted Adjustment Value	0.03	0.04
Individualistic	Percentage of Acceptances	1.92	1.10
Prosocial	Accepted Adjustment Value	0.07	0.10
Prosocial	Percentage of Acceptances	0.92	0.21
Altruistic	Accepted Adjustment Value	0.03	0.05
Altruistic	Percentage of Acceptances	0.14	0.09
Moderate	Accepted Adjustment Value	0.02	0.05
Moderate	Percentage of Acceptances	0.06	0.30

4.9.2 Persuasive Value Phrases

Figures 4.6, 4.7, 4.8, 4.9, and 4.10 give a summary of persuasive value phrases learned by our bandit models for each of the users. Due to noise during learning (other conditions such as threshold for the user given the conditions, and randomness mimicking human behavior implemented as a part of ACT-R), the results are not exact, but on the whole our model is able to successfully learn probabilities of value phrases being persuasive to users.

4.9.3 Hypothesis Results

Hypothesis 1 We can create a realistic model of human decision-making using a cognitive architecture.

Supported. Using ACT-R, we are able to create a realistic model of human decision-making, and model user responses based on internal rewards obtained by performing actions.

Hypothesis 2 Optimal spatial suggestions can be learned over time.

Supported. We are able to learn spatial thresholds of users over time, as shown in Section 4.9.1.

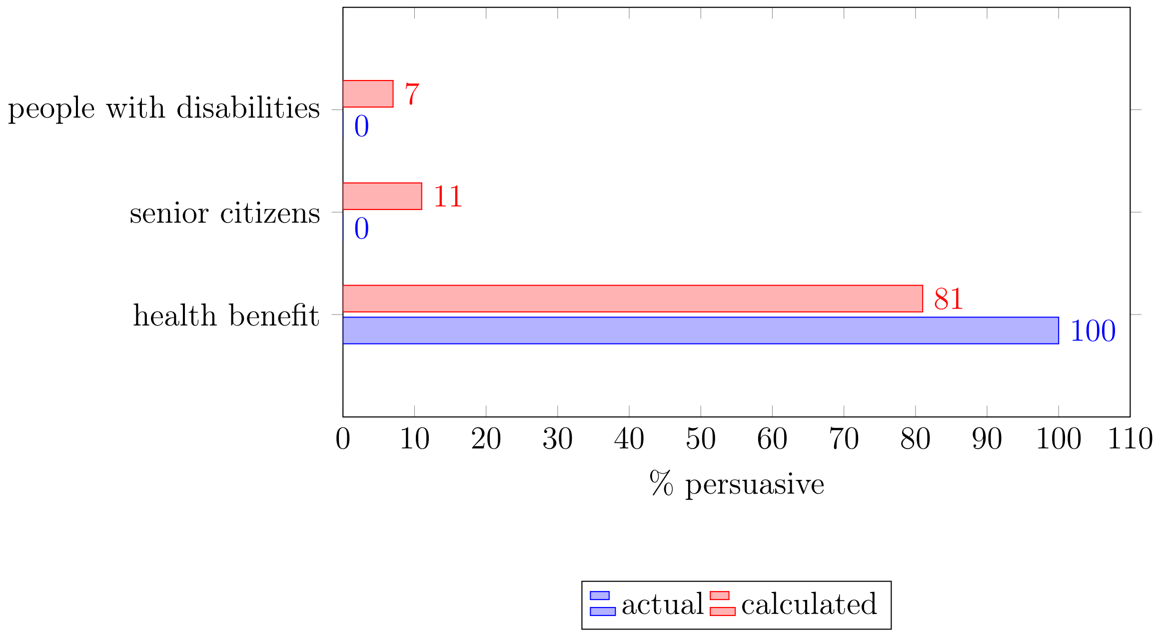


Figure 4.6: Competitive User: actual vs top three calculated persuasive value phrase percentages; the remaining 1% is divided between the rest of the phrases.

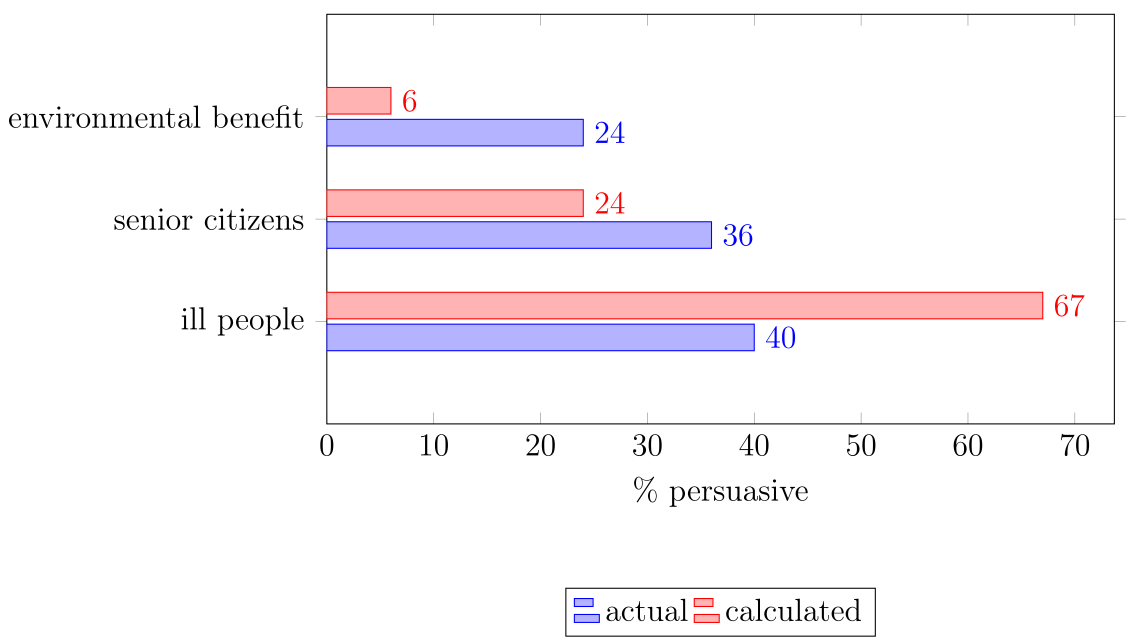


Figure 4.7: Individualistic User: actual vs top three calculated persuasive value phrase percentages; the remaining 3% is divided between the rest of the phrases.

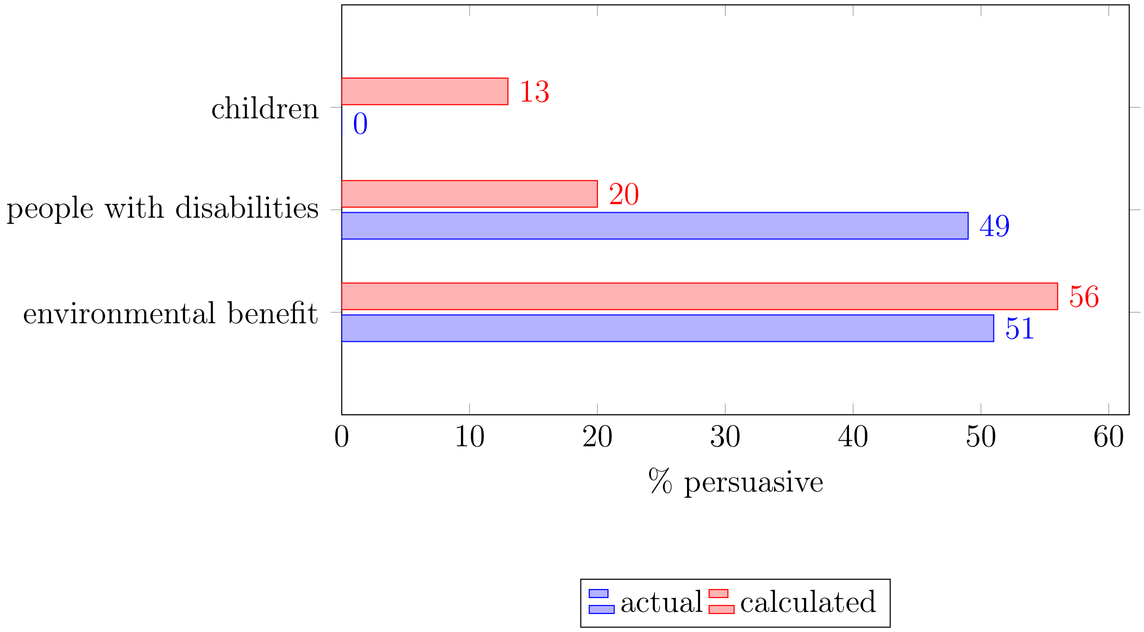


Figure 4.8: Prosocial User: actual vs top three calculated persuasive value phrase percentages; the remaining 11% is divided between the rest of the phrases.

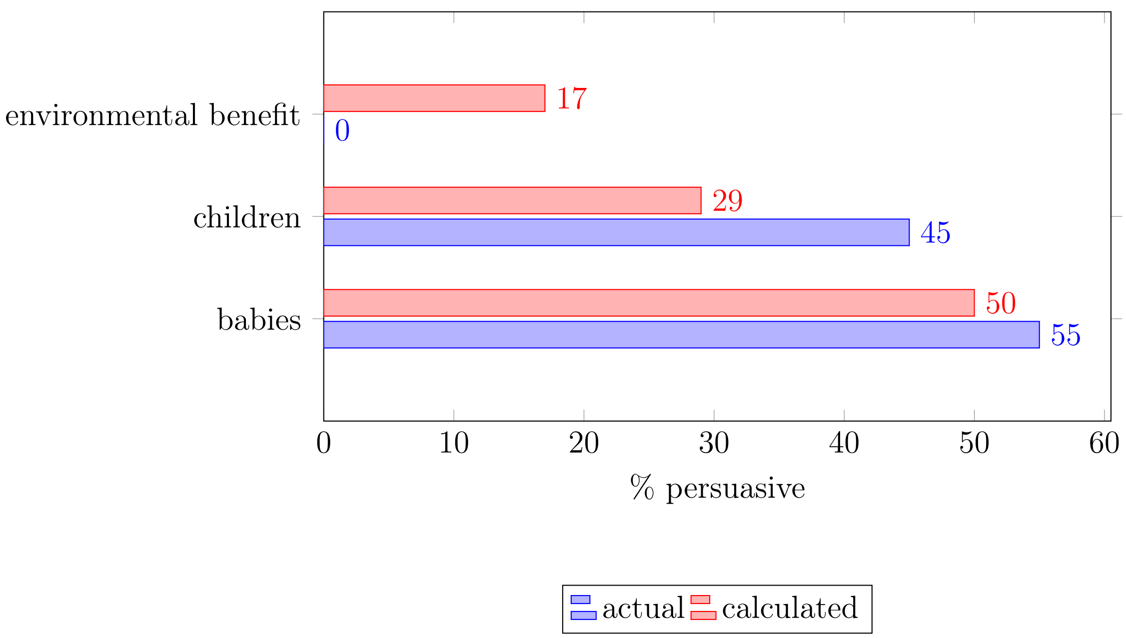


Figure 4.9: Altruistic User: actual vs top three calculated persuasive value phrase percentages; the remaining 4% is divided between the rest of the phrases.

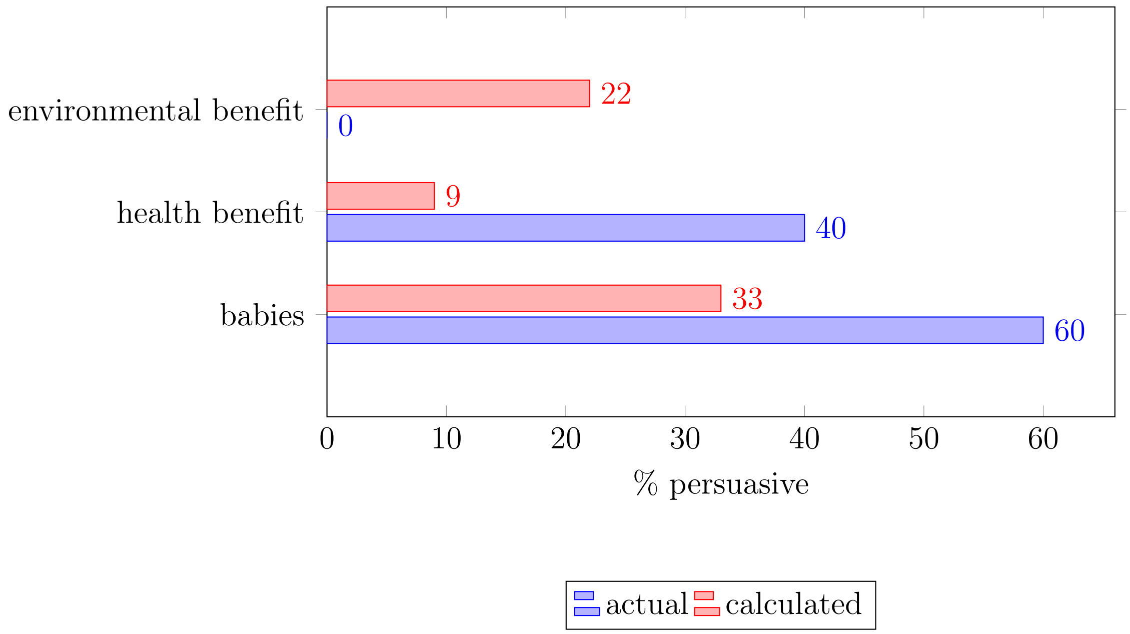


Figure 4.10: Moderate User: actual vs top three calculated persuasive value phrase percentages; the remaining 36% is divided between the rest of the phrases.

Hypothesis 3 Persuasive value phrase suggestions can be learned over time.

Supported. We are able to learn persuasive value phrases for users over time, as shown in Section 4.9.2.

Hypothesis 4 Knowledge about a user's basic profile will lead to faster learning.

Not supported. A random model converges much more quickly to the steady state than our pretrained model. However, as described in Section 4.9.1, the pretrained model proves to be, for the most part, a better starting point than an untrained model from the perspective of the user, and if provided with more or better input, it is likely to perform even better. Hence, we must compromise between a *better quality starting point* and a *quicker convergence.*

CHAPTER

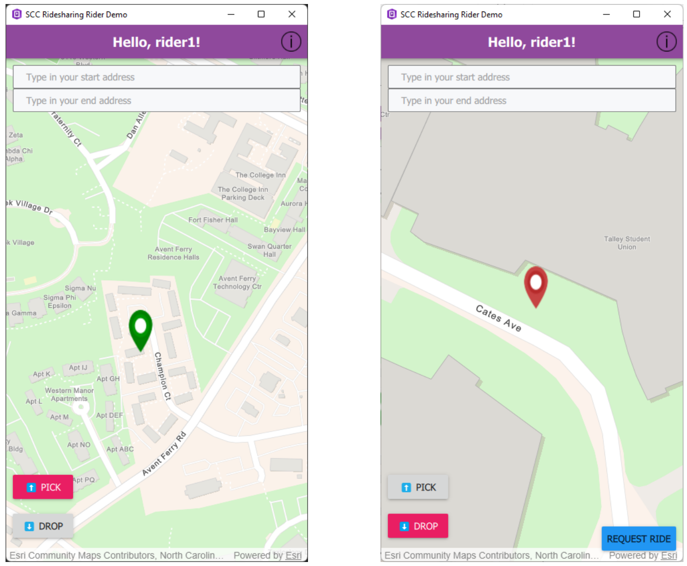
5

ILLUSTRATION VIA A MOBILE APP

To demonstrate our idea, we have built an app prototype for microtransit. We use ArcGIS, a collection of online geographic system software Booth et al. [2001] to perform the geospatial computations required to calculate candidate alternative locations. We consider multiple riders who request trips on the app, and drivers who specify their current location. Riders are clustered together based on the similarity of their routes. Considering the optimal route for the driver for each of these routes, riders are encouraged to walk to a pickup point that is closer to this route. Riders may have a disability, in which case the algorithm will not suggest any alternative pickup point.

5.1 Rider Pickup and Dropoff Locations

Riders can choose their pickup and dropoff points on the home screen after logging in, as shown in Figure 5.1. The locations can also be adjusted by moving the green or red pin, respectively. After choosing the pickup and dropoff locations, riders can request a ride by pressing the REQUEST RIDE button on the bottom right of the screen.



(a) Pickup (b) Dropoff

Figure 5.1: Rider Pickup and Dropoff Locations

5.2 Driver Location

The driver can specify their initial location as shown in Figure 5.2, which is used by the algorithm as the starting point while constructing the base route. After entering their location, drivers can check for rides by pressing the button on the bottom right of the page. Drivers are not the focus of our study but we include them for completeness of the illustration.

5.3 Suggesting an Alternative Location

We calculate the base route by considering the two farthest points in the cluster of pickup and dropoff locations and computing the route between them, considering ordered pairs of [pickup, dropoff] locations. For each rider, we compute the alternative location as the closest point on the base route from their specified location. Riders without disabilities will be suggested this alternative location, which they can accept or reject. The alternative pickup location is shown by a blue pin in Figure 5.3.

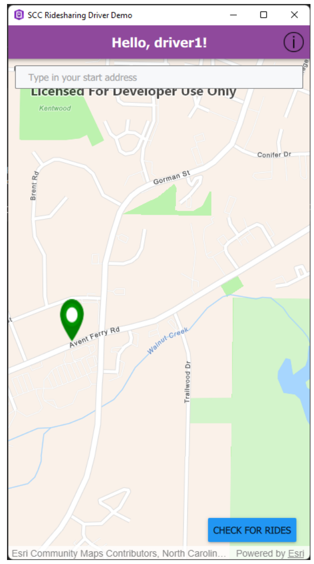


Figure 5.2: Driver Location

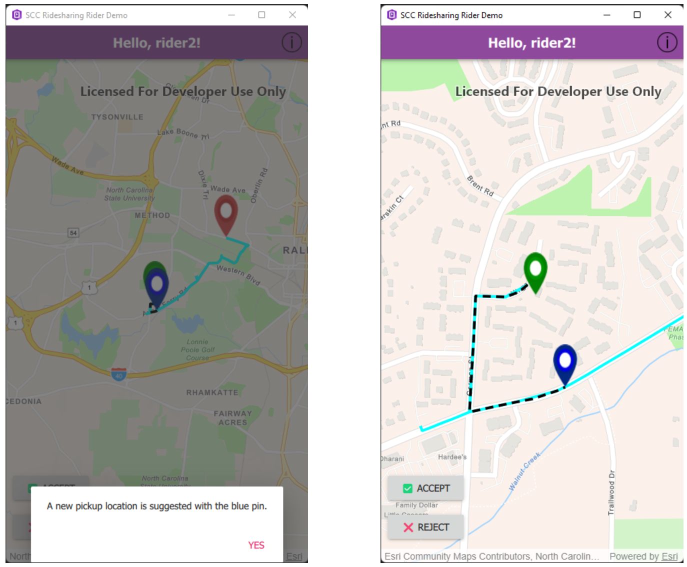
5.3.1 Presenting a Suggestion

An alternative pickup location is suggested with the blue pin as shown in Figure 5.3 (a). The rider acknowledges this before proceeding. The walking path between the original pickup point (the green pin) and the suggested location that the rider would have to take to move to the alternative point is shown in Figure 5.3 (b) as a black dotted line. Riders can choose to accept or reject the alternative location.

5.3.2 Responding to a Suggestion

If the user accepts the suggestion, the route is recalculated with the alternative pickup point, as shown in Figure 5.4 (a). A rider who accepts a suggestion obtains KARMA POINTS, which will give them a higher priority in future rides. The blue path depicts the final route to be taken by the driver.

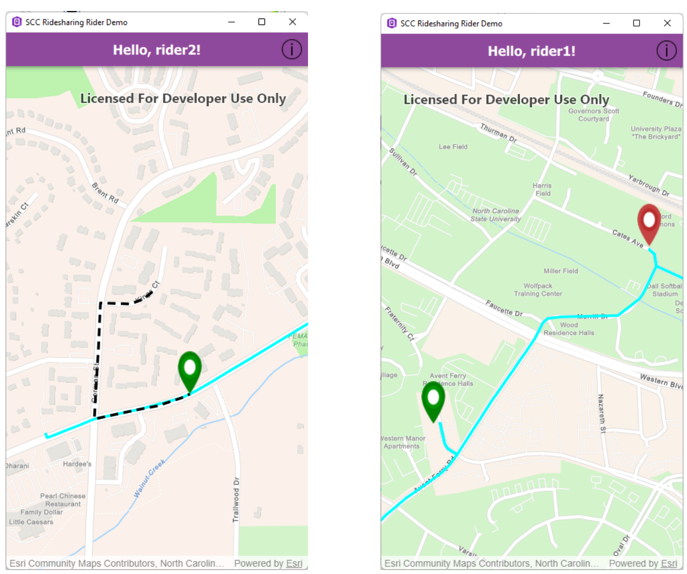
In case the suggestion is not accepted by the rider, the original path is used. As shown in Figure 5.4 (b), the route moves into the side road to pick up the rider. This would also happen in the case the rider has a disability, as in that case the rider would have to be picked up at their specified location and no suggestion will be made, since we currently only consider spatial adjustments.



(a) Alternative Point Calculated

(b) Alternative Point Details Displayed

Figure 5.3: Presenting a Suggestion



(a) Accept Suggestion

(b) Reject Suggestion

Figure 5.4: Responding to a Suggestion

CHAPTER 6

CONCLUSIONS

We present a conception of a prosocial approach to microtransit to create a more equitable and sustainable ecosystem. We demonstrate the working of our idea with a prototype app, as well as a simulation using a cognitive architecture as a surrogate for a human user, to show that user preferences can be learned with reinforcement learning and used to nudge them to help others. We show that both acceptable spatial adjustments and persuasive value phrases can be learned. We show that minimal input from the user (of their basic profile data and feature traces) is enough to pretrain a model with significant knowledge to start with, so our user is faced with minimum inconvenience even at the time of their first ride request.

With respect to the research objectives defined in Section 1.1, we show by our experiments and results in Chapter 4 that RQ1 and RQ2 are satisfied. We show that with ACT-R, we are able to model diverse users, whose varied responses to nudges indicate the differing internal satisfaction they derive from their actions. Hence we show that RQ1 (building a realistic model of human decision-making using a cognitive architecture) is met. Our results for RQ1 show that ACT-R can be used to model human decision-making in simulations where human input is required and relevant or sufficient data is not available, accounting for people with different behaviors and motivations. We show that it is possible to learn user preferences using reinforcement learning, both regarding acceptable spatial adjustments and the value phrases they are persuaded by. Thus, RQ2 (being able to learn user preferences) is satisfied. Our results for RQ2 show that reinforcement learning algorithms can be used in learning complex preferences (such as those of our temperamental users, modeled by ACT-R) as well as relatively constant preferences (persuasive value phrases, modeled to be unchanging but provided to our agent as a noisy response dependent on other conditions of our user and environment). We also maintain that knowing these preferences will help in creating a more prosocial ecosystem (RQ3). To support this hypothesis, we would require an optimizing agent that takes in user preferences while calculating the optimal adjustment for a given subproblem. Our results suggest that if we combine this work with such an optimizing agent and use it to make suggestions, it will result in a more prosocial ecosystem by suggesting adjustments that are of the least inconvenience to users while at the same time optimizing a system-level (for example, average trip time) or higher priority (for example, a more disadvantaged rider's convenience) metric.

6.1 Challenges

Using AI-based interventions to change user preferences or behavior, even for a good societal objective, is potentially ethically risky. Key challenges include a deeper understanding of consent [Singh, 2022] so that an AI agent does not violate a user's autonomy. Some researchers maintain that nudging is in itself potentially unethical [Schmidt and Engelen, 2020]. There is a worry that nudging vitiates personal autonomy. However, a nudge does not forbid any options but rather suggests or arranges options in a way that intends to *improve* people's decisions, where each person has the final say in what they choose. Nudges *steer* users toward better choices, rather than restrict them. Although nudging is meant to achieve outcomes desirable to everyone including the person being nudged, the idea can be used to manipulate people toward goals that benefit the nudger rather than the individual user. Hence, it is important to ensure that nudges are not misused.

A more general challenge is that of achieving trust. A decision about trust brings forth judgments of an agent's ability, benevolence, and integrity [Mayer et al., 1995]. The same constructs form an effective basis for assessing the trustworthiness of AI agents [Singh and Singh, 2023]. Achieving trustworthiness in the present scenario would involve care for the following kinds of concerns. For ability: the agent (and associated hardware such as vehicles) is competent and reliable, e.g., in finding safe pickup and dropoff points for riders. For benevolence: the agent acts in a user's interest, e.g., by informing them of all relevant options, including the best ones, so the user can choose freely. For integrity: the agent acts honorably, e.g., by protecting user privacy and not misusing user data. For trust, we would like not only the agent to be trustworthy but also that the users place the requisite trust in it. Besides the inherent benefits of ensuring that our STSs promote trust and that our agents are trustworthy, another motivation for trust is practical: Once a user loses trust in the system, they may elect not to participate or participate only to the extent necessary, e.g., by disregarding any nudges and thereby forgoing the prosocial outcomes we desire.

6.2 Future Work

This research opens up interesting directions for further study. First, we may explore the effects of varied ACT-R parameters, including less randomness and more shared data with the reinforcement learning agent. Here we may explore different reward functions' effects on agent performance based on the metrics defined in Section 4.9.1, and metrics that more accurately measure model performance. We may also explore the combination of the preference learning shown in this work with an optimizer to compute the resulting effect on system conditions and attest to whether it contributes to overall prosociality. We could also explore a multi-objective reinforcement learning approach for the problem, considering system benefit, user spatial preferences, and user persuasive value phrases as separate objectives to be optimized.

Second, the same idea can be applied to a temporal nudge. We can learn user preferences for adjusting their times in a comparable way, with the only difference being that the urgency of the trip is likely to be more important in temporal flexibility. Given knowledge of a user's spatial and temporal flexibility for certain conditions, we can suggest the optimal intervention to them: both which they are likely to accept and which is most beneficial for the system as a whole. If we use an optimizer algorithm for a certain cluster of riders and drivers, considering real-time traffic conditions, weather, and other environmental factors, these preferences will help decide which interventions are optimal for the given ride.

Preconceptions of users may also change based on environmental factors. For example, if it is heavily raining, people may be more sympathetic to others than they would be otherwise. If it is late, a male rider may be understanding of a female's needs and be willing to compromise. Future work can enhance the complexity of such persuasive methods.

Third, in addition to learning preferences and preconceptions of users, other methods could be used to nudge them. Gamifying the system is one way to do so [Ro et al., 2017] where people can compete for karma points or play in teams for normative influence, which would be especially effective for those who tend to conform to society. We have implemented a simple form of karma points in our app, where accepting a nudge leads to a higher position on the leaderboard or a higher preference for future requests. We could make use of choice architectures [Münscher et al., 2016] to position the nudges for maximum adjustment. Learning preferences is at the core of effective nudging, however, once those preferences are learned, there are multiple ways to increase the attractiveness of nudges to a user. Future work can also explore these areas.

REFERENCES

Evangelia Anagnostopoulou, Efthimios Bothos, Babis Magoutas, Johann Schrammel, and Gregoris Mentzas. Persuasive technologies for sustainable mobility: State of the art and emerging trends. *Sustainability*, 10(7):2128, 2018. doi: 10.3390/su10072128.

John R. Anderson and Christian Lebiere. The newell test for a theory of cognition. *Behavioral and Brain Sciences*, 26(5):587–601, 2003. doi: 10.1017/S0140525X0300013X.

John R. Anderson and Christian D. Schunn. *Implications of the ACT-R Learning Theory: No Magic Bullets*, volume 5, pages 1–33. Routledge, 2013. ISBN 9780203774526. doi: 10.4324/9780203774526.

John R. Anderson, Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin. An integrated theory of the mind. *Psychological review*, 111(4):1036–1060, 2004. doi: 10.1037/0033-295X.111.4.1036.

John Robert Anderson. ACT: A simple theory of complex cognition. *American Psychologist*, 51(4):355–365, 1996. doi: 10.1037/0003-066X.51.4.355.

Jerry Ball, Mary Freiman, Stuart Rodgers, and Christopher Myers. Toward a functional model of human language processing. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 32, 2010. URL `https://escholarship.org/uc/item/8kr28525`.

Jerry T. Ball. Advantages of ACT-R over prolog for natural language analysis. 2013. URL `http://act-r.psy.cmu.edu/wordpress/wp-content/uploads/2014/01/BRIMS2013-112.pdf`.

Daniel Balliet, Craig Parks, and Jeff Joireman. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations*, 12(4): 533–547, 2009. doi: 10.1177/1368430209105040.

Eleni Bardaka, Leila Hajibabai, and Munindar P. Singh. Reimagining ride sharing: Efficient, equitable, sustainable public microtransit. *IEEE Internet Computing (IC)*, 24 (5):38–44, September 2020. doi: 10.1109/MIC.2020.3018038.

Andreea Bobu, Marius Wiggert, Claire J. Tomlin, and Anca D. Dragan. Inducing structure in reward learning by learning features. *International Journal of Robotics Research*, 41 (5):497–518, April 2022. doi: 10.1177/02783649221078031.

Bob Booth, Andy Mitchell, et al. *Getting Started with ArcGIS*. ESRI New York, 2001.

Jelmer P. Borst and John R. Anderson. Using the ACT-R cognitive architecture in combination with fMRI data. In Birte U. Forstmann and Eric-Jan Wagenmakers, editors, *An Introduction to Model-Based Cognitive Neuroscience*, pages 339–352. Springer New York, New York, NY, 2015. ISBN 978-1-4939-2236-9. doi: 10.1007/978-1-4939-2236-9_17.

Adrian Brasoveanu and Jakub Dotlačil. *Computational Cognitive Modeling and Linguistic Theory*, chapter The ACT-R Cognitive Architecture and Its pyactr Implementation, pages 7–37. Springer International Publishing, 2020. doi: 10.1007/978-3-030-31846-8_2.

Michael D. Byrne. ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human-Computer Studies*, 55(1):41–84, 2001. doi: 10.1006/ijhc.2001.0469.

Cvetomir Dimov, Patrick H. Khader, Julian N. Marewski, and Thorsten Pachur. How to model the neurocognitive dynamics of decision making: A methodological primer with ACT-R. *Behavior Research Methods*, 52:857–880, 2020. doi: 10.3758/s13428-019-01286-2.

William G. Kennedy and Robert E. Patterson. Modeling intuitive decision making in ACT-R. In *Proceedings of the 11th International CConference on Cognitive Modeling*, volume 12, pages 1–6, 2012. URL `https://core.ac.uk/download/pdf/57706341.pdf#page=25`.

Christine Kormos, Reuven Sussman, and Bracha Rosenberg. How cities can apply behavioral science to promote public transportation use. *Behavioral Science & Policy*, 7(1):95–115, 2021. doi: 10.1353/bsp.2021.0004.

John Laird and Paul S. Rosenbloom. *The Evolution of the Soar Cognitive Architecture*. University of Michigan, Computer Science and Engineering Division, 1994.

John E. Laird. *The Soar Cognitive Architecture*. MIT press, 2019.

John E. Laird. An analysis and comparison of ACT-R and soar. *arXiv e-prints*, 2201.09305, 2022. doi: 10.48550/arXiv.2201.09305. Presented at The Ninth Advances in Cognitive Systems (ACS) Conference 2021.

Paul A. M. Van Lange. The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77(2):337, 1999. doi: 10.1037/0022-3514.77.2.337.

Christian Lebiere, Dieter Wallach, and Niels Taatgen. Implicit and explicit learning in ACT-R. In *Proceedings of the Second European CConference on Cognitive Modelling*, pages 183–189. Nottingham University Press, 1998. URL `https://research.rug.nl/en/publications/implicit-and-explicit-learning-in-act-r`.

Michael K. Martin, Cleotilde Gonzalez, and Christian Lebiere. Learning to make decisions in dynamic environments: ACT-R plays the beer game. pages 178–183, 2004. doi: 10.1184/R1/6571235.v1.

Yoshie Matsumoto, Toshio Yamagishi, Yang Li, and Toko Kiyonari. Prosocial behavior increases with age across five economic games. *PloS ONE*, 11(7):e0158671, 2016. doi: 10.1371/journal.pone.0158671.

Roger C. Mayer, James H. Davis, and F. David Schoorman. An integrative model of organizational trust. *The Academy of Management Review*, 20(3):709–734, July 1995. doi: 10.5465/amr.1995.9508080335.

Ryan O. Murphy and Kurt A. Ackermann. Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1):13–41, 2014. doi: 10.1177/1088868313501745.

Robert Münscher, Max Vetter, and Thomas Scheuerle. A review and taxonomy of choice architecture techniques. *Journal of Behavioral Decision Making*, 29(5):511–524, 2016. doi: 10.1002/bdm.1897.

Allen Newell. *Unified Theories of Cognition*. Harvard University Press, 1994.

David Peebles and Peter C.-H Cheng. Extending task analytic models of graph-based reasoning: A cognitive model of problem solving with cartesian graphs in ACT-R/PM. *Cognitive Systems Research*, 3(1):77–86, 2002. ISSN 1389-0417. doi: 10.1016/S1389-0417(01)00046-8. Computational Cognitive Modeling.

Kai Preuß. *An ACT-R model of skill development in the case of Sudoku*. PhD thesis, Albert-Ludwigs-Universität Freiburg, 2018.

Antonio Rangel. The computation and comparison of value in goal-directed choice. In Paul W. Glimcher, Colin F. Camerer, Ernst Fehr, and Russell A. Poldrack, editors, *Neuroeconomics*, chapter 28, pages 425–440. Academic Press, London, 2009. doi: 10.1016/B978-0-12-374176-9.00028-2.

Steven Ritter, John R. Anderson, Kenneth R. Koedinger, and Albert Corbett. Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14: 249–255, 2007. doi: 10.3758/BF03194060.

Michael Ro, Markus Brauer, Kathy Kuntz, Raj Shukla, and Ingo Bensch. Making cool choices for sustainability: Testing the effectiveness of a game-based approach to promoting pro-environmental behavior. *Journal of Environmental Psychology*, 53:20–30, 2017. doi: 10.1016/j.jenvp.2017.06.007.

Andreas T. Schmidt and Bart Engelen. The ethics of nudging: An overview. *Philosophy Compass*, 15(4):e12658, 2020. doi: 10.1111/phc3.12658.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *CoRR*, 1707.06347, 2017. doi: 10.48550/arXiv.1707.06347.

Susan Shaheen, Adam Cohen, Nelson Chan, and Apaar Bansal. Sharing strategies: Carsharing, shared micromobility (bikesharing and scooter sharing), transportation network companies, microtransit, and other innovative mobility modes. *Transportation*

Land Use and Environmental Planning, Berlin, Germany: Elsevier, 1:237–262, 2020. doi: 10.1016/B978-0-12-815167-9.00013-X.

Amika M. Singh and Munindar P. Singh. Wasabi: A conceptual model for trustworthy artificial intelligence. *IEEE Computer*, 56(2):20–28, February 2023. doi: 10.1109/MC.2022.3212022.

Munindar P. Singh. Norms as a basis for governing sociotechnical systems. *ACM Transactions on Intelligent Systems and Technology*, 5(1):1–23, January 2014. ISSN 2157-6904. doi: 10.1145/2542182.2542203. URL `https://doi.org/10.1145/2542182.2542203`.

Munindar P. Singh. Consent as a foundation for responsible autonomy. *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 36(11):12301–12306, February 2022. doi: 10.1609/aaai.v36i11.21494. Blue Sky Track.

Paul R. Smart and Katia Sycara. Cognitive social simulation and collective sensemaking: An approach using the ACT-R cognitive architecture. In *6th International Conference on Advanced Cognitive Technologies and Applications (COGNITIVE 2014) (24/05/14 - 28/05/14)*, May 2014. URL `https://eprints.soton.ac.uk/364297/`. Accessed 2023-05-07.

Paul R. Smart, Darren P. Richardson, Katia Sycara, and Yuqing Tang. Towards a cognitively realistic computational model of team problem solving using ACT-R agents and the ELICIT experimentation framework. In *19th International Command and Control Research Technology Symposium (ICCRTS'14) (15/06/14 - 18/06/14)*, June 2014. URL `https://eprints.soton.ac.uk/364342/`. Accessed 2023-05-07.

J. Gregory Trafton, Laura M. Hiatt, Anthony M. Harrison, Franklin P. Tamborello, Sangeet S. Khemlani, and Alan C. Schultz. ACT-R/E: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1):30–55, February 2013. doi: 10.5898/JHRI.2.1.Trafton.

Rae Zimmerman, Carlos E. Restrepo, Hannah B. Kates, and Robert Joseph. Suburban poverty, public transit, economic opportunities, and social mobility. *University Transportation Research Center*, 1, 2015. URL `https://rosap.ntl.bts.gov/view/dot/31203`.