

# Strategic Information Disclosure to People with Multiple Alternatives

Amos Azaria\* and Zinovi Rabinovich\* and Sarit Kraus\* and Claudia V. Goldman†

\* Department of Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel

† General Motors Advanced Technical Center, Herzliya 46725, Israel

azariaa1@mail.biu.ac.il, {zinovi,sarit}@cs.biu.ac.il, claudia.goldman@gm.com

## Abstract

This paper studies how automated agents can persuade humans to behave in certain ways. The motivation behind such agent's behavior resides in the utility function that the agent's designer wants to maximize and which may be different from the user's utility function. Specifically, in the strategic settings studied, the agent provides correct yet partial information about a state of the world that is unknown to the user but relevant to his decision. Persuasion games were designed to study interactions between automated players where one player sends state information to the other to persuade it to behave in a certain way. We show that this game theory based model is not sufficient to model human-agent interactions, since people tend to deviate from the rational choice. We use machine learning to model such deviation in people from this game theory based model. The agent generates a probabilistic description of the world state that maximizes its benefit and presents it to the users. The proposed model was evaluated in an extensive empirical study involving road selection tasks that differ in length, costs and congestion. Results showed that people's behavior indeed deviated significantly from the behavior predicted by the game theory based model. Moreover, the agent developed in our model performed better than an agent that followed the behavior dictated by the game-theoretical models.

## Introduction

Advanced technology allows computer systems to take an increasingly active role in people's decision-making tasks, whether as proxies for individuals or organizations (e.g., automatic bidder agents in e-commerce (Rajarshi et al. 2001)), or autonomous agents that work alongside people (e.g. training systems for diplomatic negotiation (Kraus et al. 2008)). The participants of these heterogeneous human-computer applications share common goals, but each of the participants also has its own incentives. Consider for example a centralized traffic control system that provides congestion information to commuters. The system and drivers both share the goal of getting commuters to their destination as quickly as possible. However, the system may also wish to increase the amount of tolls collected by drivers, while drivers may wish to minimize this amount.

This paper focuses on a particular class of such heterogeneous systems in which information is distributed across participants, and agents need to reason about how and when to disclose this information. There are several reasons why this task is challenging. First, the agent needs to reason about the potential effect of information disclosure on participants' possibly conflicting interests. For example, the traffic center can notify drivers that a toll-free road is vacant. It needs to reason about the effect of this notification on the toll collection for the day and on the resulting congestion for the toll-free road. Second, people's decision-making deviates from rational choice theory and is affected by a variety of social and psychological factors (Camerer 2003). For example, some people may prefer to use toll free roads, even when they are heavily congested.

The paper presents a formal model of information disclosure in a particular class of such systems. We construct a formal setting that augments an existing theoretical model of information disclosure from the literature. Participants' utilities in our setting depend on each other's actions (e.g., what information to convey to drivers, which road to choose) and the state of the world (e.g., whether the driver arrived on time, which toll road was chosen). The agent has private information that is not known to the person (e.g., the state of congestion in the different roads). The setting involves a single-shot interaction in which the agent presents true yet partial information about the state of the world to the person, and the person chooses an action based on this information. Our approach uses supervised machine learning to construct a probabilistic model of people's choices. The model constructs a utility function that depends on their own incentives as well as domain-dependent information. It accounts for the fact that people may deviate from their optimal choice given this subjective utility function. The agent integrates this model into the decision-theory framework in order to generate a probabilistic description of the state of the world and presents it to people.

We evaluated this agent in an extensive study involving more than 300 subjects participating in road selection tasks. The computer agent generated a probabilistic description of the state of the roads to present to people that maximized its interests. Our results show that (1) as shown in various different games (eg. (Cameron 1999) and (Peled, Gal, and Kraus 2011)) people's behavior significantly deviates from

that of the game theory based model; (2) the subjective utility model we constructed was able to predict people’s behavior; and (3) an agent using this model was able to lead people to make choices that were beneficial to the agent while not reducing the human participants’ benefits.

## Related Work

We are interested in scenarios where advice giving can influence the decision-making of the advice taker (see e.g., (Bonaccio and Dalal 2006) for a taxonomy). Human players participating in a coordination game were found to accept a third party’s advice, even though this third party has selfish interests in the game’s outcome (Kuang, Weber, and Dana 2007). Furthermore, communication will affect human players even if it comes from their opponents, who are directly involved in the game (see e.g., (Liebrand 1984)).

As a result, manipulative information exchange between players becomes an issue to exploit. For example, travel guidance systems have been studied for their effects on the commuting dynamics (Mahmassani and Liu 1999; Chorus, Molin, and van Wee 2006).

Manipulative interactions are most commonly captured by persuasion games (see e.g., (Kamenica and Gentzkow 2010)), where one player (Sender) possesses a key piece of information and another (Receiver) can actually act in the environment. The Sender attempts to calculate and find that portion of information which will yield maximum persuasive effect, i.e. will prompt the Receiver to choose an action most beneficial for the Sender, rather than the Receiver itself. However, to effectively manipulate the Receiver, the Sender needs to understand the motives and the decision-making process of the Receiver. Most persuasion game models fail to take into account the possibility that the Receiver’s motives may not be fully accessible to the Sender, i.e. the Receiver has private information (or type). The set of models that explicitly handle this case are called information disclosure games (Rayo and Segal 2010; Celik 2010).

## The Information Disclosure Game

The game describes an asymmetric interaction between two players: a Sender and a Receiver. Each player has a privately observed type associated with it ( $v \in V$  and  $w \in W$  respectively) that are independently sampled from commonly known distributions ( $v \sim p_V$  and  $w \sim p_W$ ). The Sender can send messages to the Receiver and the Receiver can perform actions from a set  $A$ . The utilities of the interaction between the players are given by two functions  $u_s : V \times A \rightarrow \mathbb{R}$  for the Sender, and  $u_r : V \times W \times A \rightarrow \mathbb{R}$  for the Receiver. The dynamics of such a game develop as follows:

- The Sender selects a finite set of messages  $M$ , and a disclosure rule  $\pi : V \rightarrow \Delta(M)$  that specifies the probability  $\pi(m|v)$  of sending a message  $m$  given any possible Sender’s type  $v$ . Note that  $v$  is unknown to the Sender at the time of computing this disclosure rule. We will refer to the disclosure rule as the Sender’s policy.
- The Sender declares and commits to  $(\pi, M)$ .

- The players’ private types  $v$  and  $w$  are independently sampled from  $p_V$  and  $p_W$  respectively.
- The Sender samples a message  $m \sim \pi(\cdot|v)$  and sends it to the Receiver.
- Given the message  $m$ , the Receiver performs a Bayesian update to calculate  $p_V^m \propto \pi(m|\cdot)^T \circ p_V$  where “ $\circ$ ” denotes the entry-wise product (Horn and Johnson 1991).
- Based on  $p_V^m$  and  $w$  the Receiver selects an action  $a \in A$ .
- Players obtain their respective utilities  $u_s(v, a)$  and  $u_r(v, w, a)$ .

## Solving Information Disclosure Games

To solve the information disclosure game we represent it as a mathematical program (which can be non-linear). Solving such a problem consists of optimizing the expected utility of the Sender by using a particular protocol that chooses what messages to send given its type. At the same time the action selection policy of the Receiver contributes the bounding conditions of this mathematical program. In this Section, we analyze such games formally and provide a solution assuming that the Receiver is fully rational.

### Mathematical Program

Since the Sender must commit in advance to its randomized policy, we use subgame perfect (SP) Bayesian Nash equilibrium where the only choice made by the Sender is selecting the disclosure rule (we analyze the game as if a third party sends the message to the Receiver based on the disclosure rule given to him by the sender). In the SP equilibrium the Receiver’s strategy is the best response to the Sender’s policy, simplifying the equilibrium calculations (Osborne and Rubinstein 1994).

In the following we limit the general interactions in the game to sets of Sender types  $V$ , Receiver types  $W$  and Receiver actions  $A$  that are all finite (we refer to this as the finite sets assumption). Let  $p_V^b$  denote the beliefs of the Receiver about the private type of the Sender, and the Receiver will choose an optimal action

$$a^* = \arg \max_{a \in A} \mathbf{E}_{v \sim p_V^b} [u_r(v, w, a)] \quad (1)$$

The set of feasible responses can be even further limited if the disclosure rule  $\pi$  is given. By strategically constructing the rule  $\pi$ , the Sender can influence the actions chosen by the Receiver. Since the Sender has only partial knowledge of the private value  $w$  of the Receiver, the Sender can only compute a prediction of  $a_r^*$ ,  $p_A : \Delta(V) \rightarrow \Delta(A)$ , the action choice. We denote  $p_A^m = p_A(\cdot|p_V^m)$ . Having precomputed the response function  $p_A$  of the Receiver, the Sender can calculate the expected utility of a specific disclosure rule  $\pi$  (we removed the details of the simple mathematical manipulations).

$$\begin{aligned} U_s[\pi] &= \mathbf{E}[u_s] = \sum_{v \in V} \sum_{a \in A} u_s(v, a) p(v, a) \\ &= \sum_{v \in V} \sum_{a \in A} \sum_{m \in M} u_s(v, a) p_V(v) p_A(a|p_V^m) \pi(m|v) \end{aligned} \quad (2)$$

Since we have assumed that both  $V$  and  $M$  are finite, we can formulate the disclosure rule construction as an optimization problem over the space of stochastic policies  $\pi(m|v)$  and the message space  $M$ :

$$\pi^* = \arg \max_{M, \pi: V \rightarrow \Delta(M)} U_s[\pi]$$

The following theorem shows that if an optimal solution exists, then the set of messages selected by the Sender can be limited to the size of  $|V|$ .

**Theorem 1.** *Given an information disclosure game,  $\langle V, p_V, W, p_W, A, u_r, u_s \rangle$ , with the finite sets assumption: If there is an optimal solution  $(\pi, M)$ , then there exists an optimal solution  $(\tilde{\pi}, \tilde{M})$ , where  $|\tilde{M}| \leq |V|$ .*

The sketch of the proof is as follows. Assume  $\pi$  is an optimal solution with minimum messages. We assume by contradiction that  $\pi$  has more than  $|V|$  messages. When represented as a matrix where each row represents a message and each column represents a sender type ( $v$ ), there are more rows in  $\pi$  than columns. Therefore, there exists a linear dependence between the rows, and there is some linear combination of the rows which yields  $\vec{0}$ . Choose the row with the highest absolute coefficient in this combination and write it as a linear combination of the other rows. W.l.g assume  $m_0 = \sum_{i=1}^{|M|} \alpha_i m_i$  (note that  $\forall i, |\alpha_i| \leq 1$ ). Denote  $E[m]$  as the expected utility for the system when sending message  $m$ . We now compare  $E[m_0]$  with  $\sum_{i=1}^{|M|} \alpha_i E[m_i]$ . If  $E[m_0]$  is greater, construct a new disclosure rule  $\pi'$  by replacing  $m_0$  with  $m_0 + \sum_{i=1}^{|M|} \alpha_i m_i$  and for each  $i \geq 1$  replace  $m_i$  with  $m_i \cdot (1 - \alpha_i)$ . (We merely shift probabilities from one message to another, therefore  $\pi'$  is a valid disclosure rule). This will result with  $U_s[\pi'] \geq U_s[\pi]$  contradicting  $\pi$ 's optimality. If however  $E[m] \leq \sum_{i=1}^{|M|} \alpha_i E[m_i]$ , then a new disclosure rule  $\pi''$  which removes  $m_0$  and replaces  $m_i$  with  $m_i \cdot (1 + \alpha_i)$  will have at least one message less ( $m_0$ ) and will still be optimal, contradicting the minimality of  $\pi$ .

## Finding an Optimal Policy

Unfortunately, it is intractable to find an optimal policy by solving the maximization problem of (2), since it includes non-linear constraints of the form  $\arg \max$ . The purpose of this constraint is to take into consideration the best response of the Receiver as specified in (1). Thus we need to express this requirement on the Receiver's choice using linear constraints. Toward this goal we begin with the following intuition: The real purpose of every message is to lead the Receiver to take an action that is beneficial to the Sender. However, this action should be the best response of the Receiver, given his type and the belief he forms on the state of the world. The choice of the stochastic policy is the key for influencing the Receiver's beliefs.

We begin by generating messages for each possible Receiver's response. Note that the response will depend on the Receiver's type. Formally, we define a set of functions:  $F = \{f : W \rightarrow A\}$ .  $f$  specifies an action for each Receiver's type. For each function  $f$  we create a set of messages. From Theorem 1 we know that for an optimal policy there is a need for at most  $|V|$  messages. So, in particular, there is no need for more than  $|V|$  messages to lead to a specific behaviour that is described by a function  $f$ . Thus, we create a set  $M$  of messages such that for every  $f \in F$  we generate  $V$  messages denoted by  $m_f^j, 1 \leq j \leq |V|$ .

Using this set of messages, with a size of  $|V||F|$ , we would like to consider possible policies and choose the one that maximizes the Sender's expected utility. However, we need to focus only on policies  $\pi$  that given a message  $m_f^j$  a Receiver of type  $w$  will really choose an action  $f(w)$ . We achieve this formally by designing a set of inequalities that express this condition as follows.

First, given a message  $m \in M$ , a Receiver of type  $w \in W$  and a policy  $\pi$ , the Receiver will choose an action  $a \in A$  only if he believes his expected utility from this action is higher than his expected utility from any other action. Note that after receiving a message  $m$ , the Receiver's belief that the state of the world is  $v \in V$  is proportionate to  $p_V(v)\pi(m|v)$ . Thus, the set of constraints is

$$\forall a' \in A \sum_{v \in V} u_r(v, w, a) p_V(v) \pi(m|v) \geq \sum_{v \in V} u_r(v, w, a') p_V(v) \pi(m|v) \quad (3)$$

Focusing on a specific message  $m_f^j$ , we want to satisfy these constraints for any type  $w \in W$  and require that the chosen action will be  $f(w)$ . Putting these together after some mathematical manipulations we obtain the following constraints for  $\forall w \in W$  and  $\forall a' \in A$ :

$$\sum_{v \in V} (u_r(v, w, f(w)) - u_r(v, w, a')) p_V(v) \pi(m|v) \geq 0 \quad (4)$$

Note that there may be many functions for which we will not be able to find a policy  $\pi$  that will satisfy the required constraints. However, given such a  $\pi$  and a function  $f$  we can calculate the probability  $\pi_A(a|m_f^j)$  that an action  $a \in A$  will be chosen when the Receiver gets the message  $m_f^j$ , regardless of his type. Formally, given a set  $W' \subseteq W$ , let  $\pi_W(W') = \sum_{w \in W'} p_W(w_i)$ . Then,  $\pi_A(a|m_f^j) = \pi_W(f^{-1}(a))$ .

Putting it all together, we obtain the following optimization problem:

$$\begin{aligned} \tilde{\pi}^* = \arg \max_{\pi} & \sum_{m_f^j \in M} \sum_{v \in V} u_s(v, a) p_V(v) \pi_W(f^{-1}(a)) \pi(m_f^j|v) \\ & \text{s.t.} \\ & \forall m_f^j \in M, \forall w \in W \forall a' \in A \\ & \sum_{v \in V} (u_r(v, w, f(w)) - u_r(v, w, a')) p_V(v) \pi(m_f^j|v) \geq 0 \\ & \forall v \in V \sum_{m_f^j \in M} \pi(m_f^j|v) = 1 \text{ and } \forall m_f^j \in M, \pi(m_f^j|v) \geq 0 \end{aligned}$$

That is, given a message  $m_f^j \in M$  (for all such messages), we calculate the expected utility for the Sender  $u_s(v, a)$ . The probability that  $v$  will occur is  $p_V(v)$ . The probability that  $a$  will be chosen given our discussion above is  $\pi_W(f^{-1}(a))\pi(m_f^j|v)$ . The second and third constraints verify that  $\pi$  is an appropriate policy.

Given  $\tilde{\pi}^*$  we will repeatedly remove utility dominated messages based on their linear dependency on other rows of  $\tilde{\pi}^*$  from  $M$ .

The complexity of solving the optimization problem within the above algorithm is polynomial in  $|A|$  and  $|V|$ , but exponential in  $|W|$  since  $|F| \propto |A|^{|W|}$ . Nevertheless, we found that in practice, due to an implicit dominance relation between actions, the set of appropriate functions  $F$  is much smaller than  $|A|^{|W|}$ . Thus, we were able to generate an agent that follows this game theory approach and find its policy by solving the above maximization problem. We refer to this agent as Game Theory Based Agent (GTBA).

## People Modeling for Disclosure Games in Multi-attribute Road Selection Problems

Trying to influence people's action selection presents novel problems for the design of persuasion agents. People do not adhere to the optimal, monolithic strategies that can be derived analytically. Their decision-making process is affected by a multitude of social and psychological factors (Camerer 2003). For this reason, in addition to the theoretical analysis, we propose to model the people participating in information disclosure games and integrate that model into the formal one. We assume that the agent interacts with each person only once, thus we propose a general opponent modeling approach, i.e., when facing a specific person, the persuasion agent will use models learned from data collected on other people.

The opponent modeling is based on two assumptions of human decision-making<sup>1</sup>:

- **Subjective utility functions:** People's decision-making deviates from rational choice theory and their subjective utility is a function of a variety of factors.
- **Stochastic decision-making:** People do not choose actions that maximize their subjective utility, but rather choose actions proportional to this utility. A formal model of such decision-making has been shown in (Lee 2006; Daw et al. 2006) to be of the form:

$$\alpha_r^*(a|w, p_V^b) \propto \exp\left(\mathbf{E}_{v \sim p_V^b}[u_r(v, w, a)]\right)$$

The study of the general opponent approach and its comparison with the formal model was done in the context of Road Selection Problems, which will be described next.

### Multi-attribute Road Selection Problem

The multi-attribute road selection problem is defined as an information disclosure game  $\Gamma$  with two players: a driver

<sup>1</sup>There are other models that reflect how people integrate the advice with their own private opinion (see e.g., (Boll and Larrick 2009; Yaniv 2004)). However, we chose to follow the model which is both the simplest and has been confirmed by medical studies.

and a center. The center, playing the role of the sender, can provide the driver, playing the role of the receiver, with traffic information about road conditions. In particular, the driver needs to arrive at a meeting place in  $w$  minutes. There is a set  $H$  of  $n$  highways and roads leading to his meeting location. Each road  $h \in H$  is associated with a toll cost  $c(h)$ . There are several levels of traffic load  $L$  on the roads and a set of highway network states  $V$ . A highway network state is a vector  $\vec{v} \in V$  specifying the load of each road, i.e.,  $\vec{v} = \langle l_1, \dots, l_n \rangle$ ,  $l_i \in L$ . The traffic load yields different time durations for the trip denoted  $d(\vec{v}_h, h)$  (where  $\vec{v}_h$  denotes the traffic load on road  $h$  in state  $\vec{v}$ ). If the driver arrives at the meeting on time, he gains  $g$  dollars, however he is penalized  $e$  dollars for each minute he is late. Denote the chosen road by  $a$ . Putting this together, the driver's monetary utility is given by:  $u_r(\vec{v}, a, w) = g - \max\{d(\vec{v}_h, h) - w, 0\} \cdot e$ . The driver does not know the exact state of the highway network, but merely has a prior distribution belief  $p_V$  over  $V$ . The center however, when providing information to the driver, knows the exact state, but only has prior beliefs,  $p_W$ , on the possible meeting times,  $W$ . Once given the state, the center sends a message  $m$  to the driver which may reveal data on the traffic load of the various roads. The center's utility depends on the state and the driver's chosen road  $u_s(\vec{v}_a, a)$ . It increases with the toll road  $c(a)$  and decreases with  $a$ 's load as specified in  $\vec{v}$  (see below two examples of such utility functions). The center must decide on a disclosure rule and provide it to the driver in advance (before the center is given the state). For the center, the road selection problem is therefore: given a game  $\Gamma = \langle H, L, V, W, M, c, d, p_V, p_W, u_s, u_r \rangle$ , choose a disclosure rule which will maximize  $E[u_s]$ .

Since the center's utility depends on the driver, we present a method for predicting human responses given a publicized policy and a specific message they received.

### Non-monetary Utility Estimation

Given a game  $\Gamma = \langle H, L, V, W, M, c, d, p_V, p_W, u_s, u_r \rangle$  we assume the driver chooses the road based on a non-monetary subjective utility function, denoted  $\bar{u}_r^\Gamma$  (here and in the functions defined below, we omit  $\Gamma$  when it is clear from the context). We further assume that  $\bar{u}_r$  is a linear combination of three parameters given the chosen road: travel time, road load and the toll of the road. We associate different weights ( $\alpha$ 's) with each of these parameters:  $\alpha_d$  for the trip duration time,  $\alpha_c$  for the toll cost, and for all  $l_i \in L$ , we have  $\alpha_{l_i}$ . That is, given a game  $\Gamma$ , assuming the driver knew the highway network load  $\vec{v}$  and chose road  $a$ ,  $\bar{u}_r(\vec{v}, a) = \alpha_d \cdot d(\vec{v}, a) + \alpha_c \cdot c(a) + \alpha_{\vec{v}_a}$ . Note that the utility associated with a given road depends only on the given road and its load and not on the load of other roads according to the state.

We assume stochastic decision-making and therefore, given  $\Gamma$ , we assume the driver chooses road  $h$  with a probability of  $p(a = h|\Gamma, \vec{v}) = \frac{e^{\bar{u}_r(\vec{v}, h)}}{\sum_{h' \in H} e^{\bar{u}_r(\vec{v}, h')}}$ . However when choosing an action, the driver does not know  $\vec{v}$  but only  $m$ .

Thus, the probability of choosing a road  $h$  is:

$$p(a = h|\Gamma, m) = \frac{e^{E[\bar{u}_r(\cdot, h|m)]}}{\sum_{h' \in H} e^{E[\bar{u}_r(\cdot, h'|m)]}}$$

Consider a set of games  $\mathcal{G}$  such that they all have the same set of levels of traffic load.

In order to learn the weights of the subjective utility function associated with  $\mathcal{G}$ , we assume that a set of training data  $\Psi$  is given. The examples in  $\Psi$  consist of tuples  $(\Gamma^i, m, a)$  specifying that a subject playing the driver role in the game  $\Gamma^i \in \mathcal{G}$  chose road  $a \in H$  after receiving the message  $m \in M$ . We further assume that there is a predefined threshold  $\tau > 0$  and for each  $m$  that appears in  $\Psi$  there are at least  $\tau$  examples. Denote by  $prop(\Gamma^i, m, a)$  the fraction of examples in  $\Psi$  of subjects who, when playing  $\Gamma^i$  and receiving message  $m$ , chose road  $a$ .

Next, given  $\Psi$  we aim to find appropriate  $\alpha$ s that minimize the mean square error between the prediction and the actual distribution of the actions given in the set of examples  $\Psi$ . Note that we propose to learn  $\alpha$ s across all the games in  $\mathcal{G}$ . Formally we search for  $\alpha$ s that minimize  $\sum_{\Gamma^i, m, h} (p(a = h|\Gamma^i, m) - prop(\Gamma^i, m, h))^2$ .

One may notice that the subjective utility function that we propose does *not* depend on the meeting time  $w$ . This is because the meeting time  $w$  is a private value of the driver and therefore is not specified in the examples in  $\Psi$ . However, since we are interested in the expected overall response per message of the whole population and not in predicting each individual response, if the distribution of the meeting time is left unchanged, dependence on the meeting time is embedded in the utility results. (We actually learn  $p_A^m$  directly and therefore don't depend on  $w$ ).

Next, given a specific  $\Gamma$  we incorporate the learned function  $p(a = h|m)$  as an instantiation of  $p_A^m$  into the calculation of the expected utility of a disclosure rule:

$$U_s[\pi] = \sum_{\vec{v} \in V} \sum_{h \in H} \sum_{m \in M} u_s(\vec{v}, h) p_V(\vec{v}) p(h|m).$$

Unfortunately, it means that  $U_s[\pi]$  has a very non-trivial shape (involving positive and negative exponential and polynomial expressions of its argument), and even such properties as convexity were hard to verify analytically. As a result, we chose to use the standard pattern search algorithm to find a reasonable approximation of the optimal disclosure rule with respect to  $U_s[\pi]$ . We will call this approach Opponent Model Based Agent (OMBA).

## Experimental Evaluation

In the experiments, subjects were asked to play one of two variations of the multi-attribute road selection game. Each subject played only once. All of our experiments were run using Amazon's Mechanical Turk service (AMT) (Amazon 2010)<sup>2</sup>. Participation in our study consisted of 308 subjects from the USA: 177 females and 131 males. The subjects' ages ranged from 18 to 68, with a mean of 36.

<sup>2</sup>For a comparison between AMT and other recruitment methods see (Paolacci, Chandler, and Ipeirotis 2010)

Since the experiment was based on a single multiple-choice question, we were concerned that subjects might not truly attempt to find a good solution. Therefore we only selected workers with a good reputation; they were required to pass a test before starting; and they received relatively high bonuses proportionate to the monetary utility they gained. We intended to remove all answers produced quicker than 10 seconds as being unreasonably fast. However, the average time to solve our task was greater than 1.5 minutes, and the fastest response took 23 seconds. We concluded that the subjects have considered our tasks seriously.

Our experiments aimed at answering three questions:

1. How well did the game theory based agent that finds the optimal policy of the information disclosure game, assuming that people choose the best response according to  $u_r$  (GTBA), do against people?
2. How good is OMBA at predicting drivers' road choice?
3. Does OMBA improve the center's results in comparison to the optimal policy?

## Experimental Design

We consider two variations of the multi-attribute road selection game. The first one was used for answering the first two questions and to collect data for the opponent modeling procedure. The second game was used for answering the third question, using the collected data of the first game as the training data set.

In the first game,  $\Gamma^1$ , the players had to choose one of three roads: a toll free road, a \$4 toll road and an \$8 toll road (i.e.  $H = \{h_1, h_2, h_3\}$ ,  $c(h_1) = 0$ ,  $c(h_2) = 4$  and  $c(h_3) = 8$ ). Each road could either have flowing traffic which would result in a 3 minute ride, heavy traffic which would take 9 minutes of travel time or a traffic jam which would cause the ride to take 18 minutes. That is,  $L = \{flowing, heavy, jam\}$ , and  $d(h_i, flowing) = 3$ ,  $d(h_i, heavy) = 9$  and  $d(h_i, jam) = 18$ , for all  $h_i \in H$ . An example of a state  $v$  could be  $\langle heavy, flowing, flowing \rangle$  indicating that there is heavy traffic on the toll free road and traffic is flowing on the other two toll roads. Arriving on time (or earlier) yields the player a gain of \$23 and he will be penalized \$1 for every minute that he is late. Finally, the meeting could take place in either 3, 6, 9, 12 or 15 minutes, i.e.,  $W = \{3, 6, 9, 12, 15\}$ . Thus  $u_r(\vec{v}, a, w) = 23 - \max\{d(a, \vec{v}) - w, 0\} \cdot 1$ . The prior probabilities over  $V$  and  $W$  were uniform.

The center's utility was as follows: if the subject took the toll free road, the center received \$0 regardless of the state. If the subject took the \$4 toll road, the center received \$4 if the traffic was flowing, \$2 if there was heavy traffic and \$0 if there was a traffic jam. If the subject took the \$8 toll road, the center received \$8 if the traffic was flowing, \$2 if there was heavy traffic and lost \$4 if there was a traffic jam.

In the second game,  $\Gamma^2$ , the meeting time was changed to be 12, 13, 14 and 15 minutes, i.e.,  $W = \{12, 13, 14, 15\}$ . The center's utility has also changed: the center received \$1 if the driver chose the most expensive road among those with the least traffic. Otherwise the center received \$0.

Table 1: Percentage of road usage by subjects

m	Prediction			Actual		
	Toll free	\$4 toll	\$8 toll	Toll free	\$4 toll	\$8 toll
$m_1$	0.9996	0.0003	0	0.9524	0.04761	0
$m_2$	0.1748	0.825	0.0003	0.375	0.625	0
$m_3$	0.1998	0.0588	0.7414	0.2727	0.0909	0.6364
$m_4$	0.9488	0.0324	0.0189	0.625	0.0833	0.2916

In both games the subjects were given the description of the game including the center’s preferences. Before starting to play, the subjects were required to answer a few questions verifying that they understood the game. For each subject the center received a state drawn randomly and sent a message using the disclosure rule described (see section on Information Disclosure Game). To support the subjects’ decision-making, we presented them with the distribution over the possible states that was calculated using the Bayesian rule given the message, the prior uniform distribution and the center’s policy. That is, the subjects were given  $p_V^M(m)$ . The subjects then selected a single road. As a motivation, the subjects received bonuses proportionate to the amount they gained in dollars. Comparisons between different means were performed using t-tests.

## Experimental Results

We first let the subjects play with the GTBA agent. This agent computes the game theory based policy of  $\Gamma^1$ , solving the maximization problem presented in the formal model section. Note that even though the complexity of solving this problem is high, we were able to find the optimal policy for the multi-attribute road selection games in reasonable time.

The policy of GTBA included 13 messages but 5 of them were generated with very low probability. Thus, from the 169 subjects that participated in the experiment, most of them (166 subjects) received one of 8 messages, and 3 of the subjects each received a different message.

The center received on average \$2.15 per driver. This result is significantly ( $p < 0.001$ ) higher than the utility the center would receive if all subjects were rational (i.e., maximizing  $u_r$ ), which, in expectation, was only \$0.56 per driver. Another deviation from full rationality was observed by the correlation between the time to the meeting and the road selection. For a fully rational player, the longer he has till the meeting the less likely he is to choose a toll road. However, this negative correlation between the time to the meeting and the road selection was as low as  $-0.05$ , suggesting the subjects almost ignored the meeting time. These two observations lead to the conclusion that humans tend to concentrate on the traffic in each road and it’s toll, but ignore the actual monetary value which supports our general opponent modeling approach.

## Comparing OMBA and GTBA

Using the settings of the second game,  $\Gamma^2$ , we ran two agents, GTBA and OMBA. We used the results obtained from the 166 subjects that played  $\Gamma^1$  as the training set data

$\Psi$  for OMBA. That is, the  $\alpha$ ’s for  $\bar{u}_r^{\Gamma^2}$  were learned from the subjects playing  $\Gamma^1$ , i.e.,  $\mathcal{G} = \{\Gamma^1\}$ . Both OMBA and GTBA generated 4 messages for  $\Gamma^2$ . 72 subjects played with OMBA and 64 with GTBA. OMBA performed significantly better ( $p < 0.0001$ ) than GTBA, gaining an average of 0.58 vs. 0.28 points per driver.

The three leftmost columns of Table 1 present the OBMA prediction of the probability that a person will choose one of the roads given the message that it received. The rightmost columns of Table 1 present the actual percentage of subjects receiving the messages that chose the specified roads. Using these data we found out that OMBA prediction was very accurate. There was a nearly perfect correlation of 0.94 between the prediction and actual percentage.

We also consider the best response to OMBA’s messages, assuming that the subjects choose the road that maximizes their expected  $u_r$  (as assumed by the GTBA). It turned out that the rational response to all four messages of OMBA would be to choose the toll free road regardless of the meeting time. However, as we expected, the subjects deviated from this expectation and as presented in Table 1, the majority of subjects did not choose the toll free road when receiving messages  $m_2$  and  $m_3$ .

We also checked the actual dollars earned by the subjects. When playing with OMBA the average gain per subject was \$19.94 and when playing with GTBA the average was slightly higher, \$20.42, but this difference was *not* significant. Furthermore, even when considering the subjects subjective utility the difference was not significant (7.89 vs 7.18). Thus we can conclude that while the center using OMBA obtains significantly higher utility than when using the GTBA, the drivers’ outcomes were not affected.

## Conclusions

In this paper we consider information disclosure games in which an agent tries to lead a person to take an action that is beneficial to the agent by providing him with truthful, but possibly partial, information relevant to the action selection. We first provide an algorithm to compute the optimal policy for information disclosure games. We observed that unsurprisingly, people do not follow the most rational response, and therefore provide an innovative machine learning based model that effectively predicts peoples behavior in these games. We integrate this model into our persuasion model in order to yield an innovative way for human behavior manipulation. Extensive empirical study in multi-attribute road selection games confirms the advantage of the proposed model.

Future work will be to study the application of the proposed method to settings similar to the multi-attribute road selection problems, such as the interaction between a travel agent and her customers.

### Acknowledgments

We acknowledge General Motors for supporting this research and thank Shira Abuhatzera for her useful comments.

### References

- Amazon. 2010. Mechanical Turk services. <http://www.mturk.com/>.
- Boll, J. B., and Larrick, R. P. 2009. Strategies for revising judgment: How (and how well) people use others' opinions. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 35(3):780–805.
- Bonaccio, S., and Dalal, R. S. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101:127–151.
- Camerer, C. F. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. Princeton University Press. chapter 2.
- Cameron, L. A. 1999. Raising the stakes in the ultimatum game: Experimental evidence from Indonesia. *Economic Inquiry* 37(1):47–59.
- Celik, L. 2010. Information unraveling revisited: Disclosure of horizontal attributes. SSRN. <http://ssrn.com/abstract=1719338>.
- Chorus, C. G.; Molin, E. J.; and van Wee, B. 2006. Travel information as an instrument to change car drivers travel choices: a literature review. *EJTIR* 6(4):335–364.
- Daw, N. D.; O'Doherty, J. P.; Dayan, P.; Seymour, B.; and Dolan, R. J. 2006. Cortical substrates for exploratory decisions in humans. *Nature* 441(15):876–879.
- Horn, R. A., and Johnson, C. R. 1991. *Topics in Matrix Analysis*. Cambridge University Press.
- Kamenica, E., and Gentzkow, M. 2010. Bayesian persuasion. Technical report, University of Chicago. under review.
- Kraus, S.; Hoz-Weiss, P.; Wilkenfeld, J.; Andersen, D. R.; and Pate, A. 2008. Resolving crises through automated bilateral negotiations. *Artificial Intelligence* 172(1):1–18.
- Kuang, X. J.; Weber, R. A.; and Dana, J. 2007. How effective is advice from interested parties? *J. of Economic Behavior and Organization* 62(4):591–604.
- Lee, D. 2006. Best to go with what you know? *Nature* 441(15):822–823.
- Liebrand, W. B. G. 1984. The effect of social motives, communication and group size on behaviour in an n-person multi-stage mixed-motive game. *European Journal of Social Psychology* 14:239–264.
- Mahmassani, H. S., and Liu, Y.-H. 1999. Dynamics of commuting decision behaviour under advanced traveller information systems. *Transportation Research Part C: Emerging Technologies* 7(2-3):91–107.
- Osborne, M. J., and Rubinstein, A. 1994. *A course in Game Theory*. MIT Press.
- Paolacci, G.; Chandler, J.; and Ipeirotis, P. G. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5(5).
- Peled, N.; Gal, Y.; and Kraus, S. 2011. A Study of Computational and Human Strategies in Revelation Games. *AAMAS-11*.
- Rajarshi, D.; Hanson, J. E.; Kephart, J. O.; and Tesauro, G. 2001. Agent-human interactions in the continuous double auction. In *Proc. of IJCAI-01*.
- Rayo, L., and Segal, I. 2010. Optimal information disclosure. *Journal of Political Economy* 118(5):949–987.
- Yaniv, I. 2004. Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes* 93:1–13.