A

Strategic Information Disclosure to People with Multiple Alternatives[1]

Amos Azaria, Bar-Ilan University, Israel
Zinovi Rabinovich[2], Independent researcher
Claudia V. Goldman, General Motors Advanced Technical Center, Israel
Sarit Kraus, Bar-Ilan University, Israel

In this paper we study automated agents which are designed to encourage humans to take some actions over others by strategically disclosing key pieces of information. To this end, we utilize the framework of persuasion games, a branch of game theory that deals with asymmetric interactions where one player (Sender) possesses more information about the world, but it is only the other player (Receiver) who can take an action. In particular, we use an extended persuasion model, where the Sender's information is imperfect and the Receiver has more than two alternative actions available. We design a computational algorithm that, from the Sender's standpoint, calculates the optimal information disclosure rule. The algorithm is parameterized by the Receiver's decision model (i.e. what choice he will make based on the information disclosed by the Sender) and can be re-tuned accordingly.

We then provide an extensive experimental study of the algorithm's performance in interactions with human Receivers. First, we consider a fully rational (in the Bayesian sense) Receiver decision model and experimentally show the efficacy of the resulting Sender's solution in a routing domain. In spite of the discrepancy in the Sender's and the Receiver's utilities from each of the Receiver's choices, our Sender agent successfully persuaded human Receivers to select an option more beneficial for the agent. Dropping the Receiver's rationality assumption, we introduce a machine learning procedure that generates a more realistic human Receiver model. We then show its significant benefit to the Sender solution by repeating our routing experiment. To complete our study, we introduce a second (supply-demand) experimental domain and, by contrasting it with the routing domain, obtain general guidelines for a Sender on how to construct a Receiver model.

ACM Reference Format:
Amos Azaria, Sarit Kraus, Claudia V. Goldman and Zinovi Rabinovich, 2013. Strategic Information Disclosure to People with Multiple Alternatives. *ACM Trans. Intell. Syst. Technol.* V, N, Article A (January YYYY), 21 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Computer systems have a major role in providing information to humans. This information may either be via the web (search engine, news etc.), GPS systems or decision support systems. This information is not always ingenuous; at times, this information may be intended to influence a user into performing some actions rather than others. In this paper we focus on scenarios in which an automated agent interacting with humans possesses greater information than them. The automated agent needs to reveal

[1]Preliminary results of this work appeared in [Azaria et al. 2011].
[2]Currently affiliated with Mobileye Vision Technologies Ltd., Israel, zinovi.rabinovich@mobileye.com.

Author's emails: azariaa1@cs.biu.ac.il, zr@zinovi.net, claudia.goldman@gm.com, sarit@cs.biu.ac.il.

DOI:http://dx.doi.org/10.1145/0000000.0000000

information to humans, thereby leading them to perform actions that are preferable to the agent.

Game theory, in particular persuasion games, are the most popular disciplines that study strategic reasoning as required by the mixed intelligent systems on which we are concentrating. In such games (e.g. [Crawford and Sobel 1982; Glazer and Rubinstein 2006; Kamenica and Gentzkow 2010]) two rational entities interact: a Sender and a Receiver. The Sender provides information and is assumed to be more knowledgeable and the Receiver performs an action based on the information received.

Some examples may be Google Maps [Google 2013a] or Waze [Google 2013b] applications: they know the possible settings that influence traffic congestion in the relevant countries and their times (e.g., morning rush hours) and have a distribution over the time it takes to drive on most of the roads. Similarly, automated travel agents have a lot of prior information on flights and the distribution over their delays.

In this paper, we extend these game-theoretical models as follows: While the agent holds private information (i.e., unknown to the user), it is also uncertain about the exact current state of the world. For example, the system may have an estimation of the congestion of traffic on different roads which may be unknown to the user. Still, the system may have only an estimation and not the exact value of traffic density at a particular time. We consider the setting of a one-shot interaction where the agent presents the user with information and the user chooses an action based on this information. The agent can present partial information about the state, however any information revealed by the agent must be true (unlike other work which consider manipulating the information presented to the user such as [Sarne et al. 2011]).The utility functions of both parties are different, but both depend on the state of the world and the action performed by the user. We model this setting as an optimization problem for the Sender and present an algorithm for solving it.

Following the game theory solution might not be the best strategy for an agent interacting with humans as (1) humans are not necessarily rational decision makers and therefore the agent needs to be able to model its user's reaction to the information it provides and plan accordingly. Moreover, (2) people are known to discount the advice they obtain from experts (see for example [Yaniv and Kleinberger 2000; Bonaccio and Dalal 2006] and [Kuang et al. 2007] for the case when the adviser has a monetary stake in the advice provided). For example, drivers may prefer not to pay a toll even if this decision will result in driving for a longer period of time on a more congested road. The system, on the other hand, may at times prefer to notify its drivers about a vacant toll-free road if the toll roads have heavy traffic. An intelligent traffic center needs to reason about the effect of its notifications on the toll collection for the day, on the resulting congestion for the toll-free road and on the user's possible reaction to the revealed information.

To face the challenge of human deviation from fully rational behavior we present the Linear weighted-Utility Quantal response (LUQ) human model which relies on the following two assumptions: Linear Weighted-Utility, i.e. people's subjective utility is a linear combination of attributes; and Logit quantal response: The probability that people will chose a certain action is proportional to the action's subjective utility.

We ran extensive evaluations involving a total of over 700 human subjects in two different domains. One considers a road selection problem (described in Section 5.1) and the second one considers a supply-demand interaction detailed in Section 5.2. We discovered that, in the road selection problem domain, people deviated from rational behavior and therefore an agent based on the LUQ method significantly outperformed a game theory-based agent. However, in the supply-demand domain, people behaved nearly rationally and thus the LUQ based agent and the game theory-based agent's performance did not differ significantly.

To summarize, our key contributions in this paper are:

— An extended persuasion game model for human-agent interaction with asymmetric information and two-sided uncertainty;
— A formal solution algorithm for the model, parameterized by the Receiver (human) behavior model;
— The LUQ method for building a human behavioral model pertinent to the Sender-Receiver type interaction;
— A methodology determining when one can assume rational behavior and thus use the game theory approach and when one should use the LUQ method.

The rest of the paper is organized as follows. Section 2 describes the necessary background to our work and positions our paper relative to other persuasion studies. Section 3 formally defines the interaction model which we consider, while Section 4 provides its theoretical analysis and an algorithmic solution. Our experimental designs and results begin with Section 5 which describes bounded rationality models of human decisions, both general and specific to our experimental domains. Section 6 describes the exact experimental setup and parameters that we used for our domains, and describes the experimental outcomes and statistics. Finally, we give some concluding remarks in Section 7.

2. RELATED WORK

In our preliminary work [Azaria et al. 2011] we considered an information disclosure game in which only one side, the Receiver, had incomplete information about the state of the world, while the Sender had full information regarding the state of the world. In this work we explicitly deviate from this assumption and force the Sender to deal with incomplete information resulting in a two-sided uncertainty game. We furthermore provide the full proofs in the theoretical section, conduct new experiments and add an extra domain which yields interesting results.

Related scenarios to our problems are settings 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 et al. 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 et al. 2006].

Game theory researchers have long studied the manipulative interactions in the context of persuasion games (see e.g. [Milgrom and Roberts 1986; Crawford and Sobel 1982]). In these games, the Sender (the highly informed player) attempts to calculate and find that portion of information which will yield the maximum persuasive effect, i.e. will prompt the Receiver (player capable of acting in the world, and whose actions determine the welfare of both players) to choose an action which is most beneficial for the Sender, rather than the Receiver itself. Though this interaction narrative is common to all works utilizing the game theoretic persuasion, its detailed formalizations vary significantly. For example, Glazer and Rubinstein [Glazer and Rubinstein 2006] and Rayo and Segal [Rayo and Segal 2010] study the case where the Receiver has only two options: either accept the action associated with the world's current state or decline it. Such would be the case, for instance, if following or skipping a sponsored search link. The search engine would play the role of the Sender and provide additional information about the sponsored link, while the browsing Receiver would consider the relevance of the link from its point of view. In contrast, Kamenica and Gentzkow [Kamenica and Gentzkow 2010], as well as our current work, view scenarios where the set of action alternatives is greater than the simple binary choice. Another parameter of a persuasion game model is the amount of knowledge that the Sender has about the world and the Receiver. For instance, Turkay Pillai [2009] assumes that the Receiver may have some private information (a *type*) and studies how the Sender can refine its knowledge of the Receiver's utility over time. In contrast, our model of the interaction is single-shot, where such a refinement is not possible. We also assume an additional impediment for the Sender, specifically that it does not have complete knowledge of the state of the world.

Interestingly, the assumption of complete information is not limited to game theoretic persuasion approaches. In a recent work in e-Commerce, Hajaj et al. [Hajaj et al. 2013] consider comparison shopping agents (CSAs). They suggest a set of methods for affecting users decisions based on selective disclosure of information, aiming to influence users not to query additional CSAs. However, Hajaj et al. do not allow the presentation of uncertain information, using probabilities as we do, but merely allow the agent do decide whether to present a certain shop or not.

Fenster et al. [Fenster et al. 2012] design an agent which influences human decision-making in a conversational setting. Their agent tries to convince the human by providing examples for her to emulate, or by providing justifications for a certain choice. The work studied an environment where the human had to select a location for a school. The agent interacted with the human and attempted to convince her to choose a certain location. However, in a striking contrast to our work, Fenster et al [Fenster et al. 2012] have no uncertainty involved.

Although game theory has provided a wide variety of persuasive models and methods, the core assumption of rationality and equilibrium choice remain. Yet, following equilibrium strategies is often less beneficial in practical applications, where agents need to interact with people (see e.g. [Hoz-Weiss et al. 2008; Peled et al. 2011; Azaria et al. 2012]). Thus, there's a need to develop persuasion solutions where an alternative (bounded-rational or learned/mined) decision model of human behavior can be easily incorporated. Indeed such interaction methods, which are capable of combining psychological factors and human decision-making theory with machine learning methods towards creating a human model, have been demonstrated to be successful. (see e.g. [Rosenfeld et al. 2012; Gal and Pfeffer 2007; Oshrat et al. 2009; Rosenfeld and Kraus 2011; Peled et al. 2011; Azaria et al. 2012a; Azaria et al. 2012b]).

3. THE INFORMATION DISCLOSURE GAME WITH TWO-SIDED UNCERTAINTY

In this section we will formally describe the protocol of the interaction between the human user and the advising agent. To this end we will use the terminology and general format of (Bayesian) persuasion games [Kamenica and Gentzkow 2010] (hence, naming the human user a Receiver, and the agent a Sender) and a guided route selection example as intuition.

The game describes an asymmetric interaction between two players: a Sender and a Receiver. The Receiver has a privately observed type associated with it ($\theta \in \Theta$) that is sampled from a commonly known distribution ($\theta \sim p_\Theta$). 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 depend on the state of the world $v \in V$ that is sampled independently from the commonly known distribution $v \sim p_V$. The Sender can obtain an observation of the state of the world $\omega \in \Omega$ that is sampled from the commonly known distribution $\omega \sim p_\Omega(\cdot|v)$. 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 $u_r : V \times \Theta \times A \rightarrow \mathbb{R}$ for the Receiver.

In our example, θ can correspond to the tolerance or patience exhibited by a driver and influence his utility (see below). The messages sent by the Sender naturally correspond to the traffic management center sending route information. The action chosen by the Receiver corresponds to the driver choosing a specific route. The state of the world corresponds to different traffic conditions across the road network with an appropriate statistic. The traffic management center can monitor the traffic conditions with some degree of uncertainty. The utility functions in our example scenario describe how content the user would be (u_r) if he took a specific route ($a \in A$) given his patience ($\theta \in \Theta$) and current traffic conditions ($v \in V$), and respectively (u_s) how profitable it would be for the traffic management center if the driver adopted a particular route ($a \in A$) given the current traffic conditions ($v \in V$).

The game unfolds as follows:

— The Sender selects a finite set of messages, M, and a disclosure rule $\pi : \Omega \rightarrow \Delta(M)$, where $\Delta(\cdot)$ denotes the space of all distributions over a set. In other words, the disclosure rule specifies the probability $\pi(m|\omega)$ of sending a message m given any possible Sender's observation ω. Note that v is unknown (even through observation) 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 computes the *effective disclosure rule* $\pi_\Omega(m|v) = \sum_{\omega \in \Omega} \pi(m|\omega) p_\Omega(\omega|v)$.
— The Sender declares and commits to (π_Ω, M).[3]
— The Receiver's private types θ and the state of the world v are independently sampled from p_Θ and p_V, respectively.
— The Sender is supplied with the observation $\omega \sim p_\Omega(\cdot|v)$.
— The Sender samples a message $m \sim \pi(\cdot|\omega)$ and sends it to the Receiver.
— Given the message m, the Receiver performs a Bayesian update to calculate $p_V^m \propto \pi_\Omega(m|\cdot)^T \circ p_V$, where "$\circ$" denotes the entry-wise product [Horn and Johnson 1991].
— Based on p_V^m and θ the Receiver selects an action $a \in A$.
— Players obtain their respective utilities $u_s(v, a)$ and $u_r(v, \theta, a)$.

4. SOLVING INFORMATION DISCLOSURE GAMES WITH TWO-SIDED UNCERTAINTY

To solve the information disclosure game we represent it as a mathematical program (which can be non-linear). Solving such a problem consists of maximizing the expected utility of the Sender by using a particular protocol that chooses what messages to send given its observation of the state of the world. 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.

4.1. Mathematical Program

Since the Sender must commit in advance to its randomized policy, we use a 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].

We limit the possible states of the world V, the Receiver types Θ, the set of observations Ω and the Receiver actions A to finite sets (we refer to this as the finite sets

[3]In our route selection scenario the above stages correspond to the traffic management center describing and advertising its services.

assumption). Let p_V^b denote the beliefs of the Receiver about the state of the world. The Receiver will choose an optimal action:

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

The set of feasible responses can be limited even further if the disclosure rule π is given. By strategically constructing the rule π, the Sender can influence the actions chosen by the Receiver. Since the Sender has only partial knowledge of the private value θ 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 π (we removed the details of the simple mathematical manipulations).

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

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

$$\pi^* = \arg\max\nolimits_{M, \pi : V \rightarrow \Delta(M)} U_s[\pi] \tag{2}$$

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 $|\Omega|$.

THEOREM 4.1. *Given an information disclosure game, $\langle V, p_V, \Theta, p_\theta, \Omega, p_\Omega, A, u_r, u_s \rangle$, with the finite sets assumption (i.e. V, Ω and A are finite). If there is an optimal solution (π, M) where $|M| < \infty$, then there exists an optimal solution $(\widetilde{\pi}, \widetilde{M})$, where $|\widetilde{M}| \leq |\Omega|$.*

Theorem 4.1 shows that an optimal solution with a finite message space can be transformed so that the set of messages does not exceed $|\Omega|$. However, it is possible to question whether an optimal solution with a finite message set in fact exists. The following theorem deals with that question, demonstrating that a countable set of messages of an optimal solution can always be replaced by a finite set.

THEOREM 4.2. *Given an information disclosure game, $\langle V, p_V, \Theta, p_\theta, \Omega, p_\Omega, A, u_r, u_s \rangle$, with the finite sets assumption (i.e. V, Ω and A are finite). If the optimal U_s is attainable by some protocol (π, M), then there is an optimal solution with a finite message space.*

We give the complete proofs of Theorems 4.1 and 4.2 in Appendix A rather than here due to their technicality. Their intuition, however, is easily outlined. For Theorem 4.1, we show that the effects induced by the extra messages can be achieved by distributing the information that they transfer to other messages without effecting the Sender's utility. The re-distribution process relies on the linear properties of the disclosure rule as a matrix. In turn, for Theorem 4.2, we show that the utility gains obtained from almost all, but a finite number, of messages is negligible and so is the information which they provide to the Receiver. In fact, they can be aggregated into a single message (thus reducing the total number of used messages to finite) without impacting the Sender's utility.

4.2. Finding an Optimal Policy

Unfortunately, directly finding an optimal policy by solving the disclosure rule maximization problem presented in Equation 2 is intractable, since it includes a strongly non-linear component. More specifically, it assumes availability of the Receiver's best response (defined by Equation 1) in a (closed) functional form. However, it is possible to circumvent this hindrance. Instead of assuming a functional best response form, we expand Equation 2 by a set of constraints that compare the Receiver's utility from its chosen action to that of all other actions available to him/her. In other words, we transform an explicit (functional) non-linear representation of the Receiver's response into an implicit (constraints-based) linear form.

We begin by generating messages for each possible response from the Receiver. Note that the response will depend on the Receiver's type. Formally, we define a set of functions: $F = \{f : \Theta \rightarrow A\}$. f specifies an action for each Receiver's type. For each function f we create a set of messages. From Theorem 4.1 we know that for an optimal policy there is a need for at most $|\Omega|$ messages. Therefore, there is no need for more than $|\Omega|$ messages to lead to a specific behavior that is described by a function f. Thus, we create a set M of messages such that, for every $f \in F$, we generate Ω messages denoted by m_f^j, $1 \leq j \leq |\Omega|$.

Using this set of messages with a size of $|\Omega||F|$, 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 π where, given a message m_f^j, a Receiver of type θ will really choose an action $f(\theta)$. 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 $\theta \in \Theta$ and a policy π_Ω, the Receiver will choose an action $a \in A$ only if he believes that 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 $p_V(v)\pi_\Omega(m|v)$. Thus, the set of constraints is

$$\forall a' \in A \quad \sum_{v \in V} u_r(v, \theta, a) p_V(v) \pi_\Omega(m|v) \geq \sum_{v \in V} u_r(v, \theta, a') p_V(v) \pi_\Omega(m|v) \tag{3}$$

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

$$\sum_{v \in V} (u_r(v,\ f(\theta)) - u_r(v, \theta, a')) p_V(v) \pi_\Omega(m|v) \geq 0 \tag{4}$$

Note that there may be many functions for which we will not be able to find an effective policy π_Ω that will satisfy the required constraints. However, given such a π_Ω 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 $\Theta' \subseteq \Theta$, let $\pi_\Theta(\Theta') = \sum_{\theta \in \Theta'} p_\Theta(\theta_i)$. Then, $\pi_A(a|m_f^j) = \pi_\Theta(f^{-1}(a))$.

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

$$\tilde{\pi}^* = \arg\max_{\pi} \sum_{m_f^j \in M} \sum_{\substack{a \in A \\ v \in V}} u_s(v, a) p_V(v) \pi_\Theta(f^{-1}(a)) \pi_\Omega(m_f^j|v)$$

$$s.t.$$

$$\pi_\Omega = \pi p_\Omega$$

$$\forall m_f^j \in M, \forall\ \theta \in \Theta\ \forall a' \in A$$

$$\sum_{v \in V} (u_r(v, \Theta, f(\Theta)) - u_r(v, \theta, a'))p_V(v)\pi_\Omega(m_f^j|v) \geq 0$$

$$\forall \omega \in \Omega \sum_{m_f^j \in M} \pi(m_f^j|\omega) = 1$$

$$\forall m_f^j \in M \ \pi(m_f^j|\omega) \geq 0$$

The complexity of solving the optimization problem within the above algorithm is polynomial in $|A|$, $|V|$ and $|\Omega|$, but exponential in $|\Theta|$ since $|F| \propto |A|^{|\Theta|}$.

We refer to this agent as the Game Theory Based Agent (GTBA).

5. PEOPLE MODELING FOR DISCLOSURE GAMES IN MULTI-ATTRIBUTE SELECTION PROBLEMS

Trying to influence people's action selection presents novel problems for the design of persuasion agents. People often 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 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 from other people.

The opponent modeling is based on two assumptions of human decision-making:

— **Linear Weighted-Utility:** People's decision-making deviates from rational choice theory; they use a subjective utility function which is a linear combination of a set of attributes. This utility function may divert from the expected monetary utility function.

— **Logit quantal response (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:

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

We name this method for human modeling: Linear weighted-Utility Quantal response (LUQ). (This method was also proved to be successful in modeling human behavior in security games [Nguyen et al. 2013].) The study of the general opponent approach and its comparison with the formal model was done in the context of two games. The Multi-attribute Road Selection Problem with two-sided uncertainty about road traffic and the Sandwich Game with two-sided uncertainty regarding the number of attendees of a particular event. Next we will describe the two games and explain their differences.

5.1. Multi-attribute Road Selection Problem with Two Sided Uncertainty

The multi-attribute road selection problem with two-sided uncertainty about the state of the world is an extension of the game that was studied in [Azaria et al. 2011]. It is defined as an information disclosure game Γ_ρ with two players: a driver 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 θ 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., $\vec{v} =< l_1, ..., l_n >$, $l_i \in L$. The traffic load yields a different time duration 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:

$$u_r(\vec{v}, a, \theta) = g - max\{d(a, \vec{v}) - \theta, 0\} \cdot e \tag{5}$$

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 also does not know what the exact state of the highway network will be when the driver drives along the chosen road (e.g., even though the traffic flows on a given road, an accident can occur shortly and the road will be blocked). However, given its observations, the center has a better estimation of the state of the roads. The center has only prior beliefs, p_Θ, regarding the possible meeting times, Θ. Once given the observations on the state, the center sends a message m to the driver which may reveal data about the traffic load of the various roads. The center's utility depends on the actual traffic load 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 some information on the road loads). For the center, the road selection problem is therefore: given a game $\Gamma = \langle H, L, V, \Omega, M, c, d, p_V, p_\Omega, u_s, u_r \rangle$, choose a disclosure rule which will maximize $E[u_s]$.

5.2. The Sandwich Game

The Sandwich Game is defined as an information disclosure game Γ_σ with two players: a seller and an organizer. The organizer, playing the role of the Sender, can provide the seller, playing the role of the Receiver, with information regarding the anticipated conference attendees. The organizer himself receives noisy information regarding the exact number of attendees (can be interpreted as the number of people who registered to the conference during the pre-conference registration). The seller must decide in advance how many sandwiches to prepare for the conference (a). The sandwiches are sold for a fixed price $\bar{c}$, and it is assumed that each conference attendee buys a single sandwich. Each seller is associated with a private type θ which indicates the cost for preparing each possible number of sandwiches. Thus the seller's monetary utility given the number of attendees (v), the number of sandwiches prepared (a) and θ is given by $u_r(v, \theta, a) = \min\{a, v\} \cdot \bar{c} - \theta(a)$. Depending on the actual conference size, the organizer is assumed to have some preferences as to the number of sandwiches prepared by the seller ($u_s(v, a)$).

5.3. Hypothesis

In the original Road Selection problem presented in [Azaria et al. 2011], which considered only one-sided uncertainty, the agent using the general opponent modeling approach achieved a significantly higher utility than the GTBA agent. The major cause for this effect is that people preferred not to choose jammed roads in the game even when they could be on time to their meetings and therefore did not attempt to maximize their monetary values. Thus, we hypothesize that a similar agent (relying on the LUQ method for human modeling) for the two-sided uncertainty Road Selection problem will also outperform the GTBA agent in the extended game. In addition we designed the Sandwich Game, a new game in which the goal of the players is to maximize their monetary values. We expect that, in such situations, people's behavior will be more motivated to maximize their expected monetary values and GTBA may perform similar to an agent which relies on the LUQ method for human modeling.

5.4. Non-monetary Utility Estimation for the Road Selection Problem with Two-Sided Uncertainty

Given a game $\Gamma_\rho =< H, L, V, \Theta, M, c, d, p_V, p_\Theta, u_s, u_r >$, based on the LUQ method for human modeling, we assume that the driver chooses the road based on a non-monetary subjective utility function, denoted $\bar{u}^{\Gamma_\rho}$ (here and in the functions defined below, we omit Γ_ρ when it is clear from the context). We further assume that $\bar{u}$ 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 (αs) with each of these parameters: α_d for the trip duration time, α_c for the toll cost, and for all $l_i \in L$ we have α_{l_i}. That is, given a game Γ_ρ, assuming that the driver knew the highway network load $\vec{v}$ and chose road a:

$$\bar{u}_\rho(\vec{v}, a) = \alpha_d \cdot d(\vec{v}, a) + \alpha_c \cdot c(a) + \alpha_{\vec{v}_a} \quad (6)$$

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 that the user uses logit quantal response and therefore, given Γ_ρ, we assume that the driver chooses road h with a probability of

$$p(a = h|\Gamma_\rho, \vec{v}) = \frac{e^{\lambda \bar{u}_\rho(\vec{v}, h)}}{\sum\limits_{h' \in H} e^{\lambda \bar{u}_\rho(\vec{v}, h')}} \quad (7)$$

where λ is a parameter. However, since $\bar{u}_\rho(\vec{v}, h)$ has an extra degree of freedom, we set $\lambda = 1$.

When choosing an action, the driver does not know $\vec{v}$ but only m. Thus, the probability of choosing a road h is:

$$p(a = h|\Gamma_\rho, m) = \frac{e^{E[\bar{u}_\rho(\cdot, h|m)]}}{\sum\limits_{h' \in H} e^{E[\bar{u}_\rho(\cdot, h'|m)]}}$$

Consider a set of games $\mathcal{G}_\rho$ 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 $\mathcal{G}_\rho$, we assume that a set of training data Ψ is given. The examples in Ψ consist of tuples (Γ^i_ρ, m, a) specifying that a subject playing the driver's role in the game $\Gamma^i_\rho \in \mathcal{G}_\rho$ 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 Ψ there are at least τ examples. Denote by $prop(\Gamma^i_\rho, m, a)$ the fraction of examples in Ψ of subjects who, when playing Γ^i_ρ and receiving message m, chose road a.

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

One may notice that the subjective utility function that we propose does *not* depend on the meeting time θ. This is because the meeting time θ is a private value of the driver and therefore is not specified in the examples in Ψ. 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^m_A directly and therefore don't depend on θ).

Next, given a specific Γ_ρ, 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:

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

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 in order to find a reasonable approximation of the optimal disclosure rule with respect to $U_s[\pi]$.

5.5. Non-monetary Utility Estimation for the Sandwich Game with Two-Sided Uncertainty

Based on the LUQ method for human modeling we assume that the seller decides on the number of sandwiches to prepare based on the following subjective utility function: $\alpha_1 \cdot \min\{a, v\} + \alpha_2 \cdot \max\{(a - v), 0\}$. That is, the seller tries to maximize the number of sandwiches sold, and minimize the number of sandwiches thrown away (we anticipate that α_2 will be negative). For similar reasons to those mentioned in Section 5.4, the proposed subjective utility function does not depend on θ. According to LUQ we assume logit quantal response. Learning these αs and building an optimal policy is conducted in a method identical to that of the road selection problem. Each of these proposed agents which rely on the LUQ method (for each of the two domains) will be called LUQ Agent (LUQA).

6. EXPERIMENTAL EVALUATION

Our experiments are 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?
(2) Does LUQA improve the Sender's results in comparison to GTBA?
(3) Do the answers to the above questions depend on the domain and, if so, given a domain, can we provide a way to predict whether LUQA or GTBA will perform better?

6.1. Experimental Design

In both games the subjects were given the description of the game including the Sender'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 in section 3. 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 action (either a number of sandwiches to prepare or a 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.

We considered two variations for each of the two games (the sandwich game and the road selection game). The first one was used for answering the first question and in order to collect data for the opponent modeling procedure. The second variation was used for answering the second question, using the collected data of the first variation as the training data set. We now describe the parameters used for both variations of the sandwich game and the road selection game.

Table I. Seller types

Number of sandwiches	Cost for Type 1	Cost for Type 2	Cost for Type 3
None	\$0	\$0	\$0
10	\$5	\$8	\$12
20	\$9	\$12	\$15
30	\$14	\$16	\$18
50	\$20	\$20	\$20

6.1.1. Road Selection Game. In the first game, Γ^1, the players had to choose one of three roads: a toll free road, a \$4 toll road or 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., $\Theta = \{3, 6, 9, 12, 15\}$. Thus $u_r(\vec{v}, a, \theta) = 23 - max\{d(a, \vec{v}) - w, 0\} \cdot 1$. 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, Γ^2, the meeting time was changed to be in $12, 13, 14$ and 15 minutes, i.e., $\Theta = \{12, 13, 14, 15\}$. The center's utility was 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.

6.1.2. Sandwich Game. The conference size (v) had either no participants (a canceled conference), 20 participants (a small conference), 30 participants (a medium conference), 40 participants (a large conference) or 50 participants (a huge conference).

The number of sandwiches prepared by the seller (a) was in $\{0, 20, 30, 40, 50\}$ as well. Recall that the seller's utility function is given by $u_r(v, \theta, a) = \min\{a, v\} \cdot \bar{c} - \theta(a)$. We set $\bar{c}$ (the sandwich retail price) to $\$1$. We used three different private types types (θ), which indicate the cost for preparing each possible number of sandwiches. Table I shows the different private types used.

We considered two different utility functions for the organizer in the sandwich game. In the first game, Γ^1_σ, the system wanted the seller to prepare more sandwiches than needed, unless the conference had 50 attendees. In Γ^2_σ the system wanted the seller to prepare less sandwiches than needed, unless the conference was canceled (0 attendees). The utility function was chosen such that the utility for the organizer and the seller will be different and not linearly dependent. The observation table is shown in Table II. As can be seen in the table, if the organizer observes that the conference will be canceled, then in fact it will be. In any other case there is an 85% chance that the organizer will observe the correct state. Even if the state observed is incorrect, the actual state is not too far off, unless the conference is unexpectedly canceled.

6.2. Human Subjects

In the experiments, subjects were asked to play either the sandwich game or the multi-attribute road selection game with two-sided uncertainty. As mentioned above, each of

Table II. Observation table in the sandwich game

Observation	Probability for actual conference size				
	Canceled	Small	Medium	Large	Huge
Canceled	1	0	0	0	0
Small	0.08	0.85	0.06	0.01	0
Medium	0.02	0.06	0.85	0.06	0.01
Large	0.01	0.03	0.06	0.85	0.05
Huge	0.06	0	0.03	0.06	0.85

the games had two different variations which differed in the system utility function. Each subject played only once. All of our experiments were run using Amazon's Mechanical Turk service (AMT) [Amazon 2013][4]. Participation in our study consisted of 713 subjects from the USA: 56.2% females and 43.8% males. The subjects' ages ranged from 18 to 74, with a mean of 34 and a standard deviation of 11.3. The subjects participated in the following experiments:

— 173 subjects participated in Γ^1_ρ, which is the first game played in the road selection game, using GTBA.
— 102 subjects participated in Γ^2_ρ, which is the second game played in the road selection game, using GTBA.
— 119 subjects participated in Γ^2_ρ, which is the second game played in the road selection game, using LUQA.
— 100 subjects participated in Γ^1_σ, which is the first game played in the sandwich game, using GTBA.
— 106 subjects participated in Γ^2_σ, which is the second game played in the sandwich game, using GTBA.
— 113 subjects participated in Γ^2_σ, which is the second game played in the sandwich game, using LUQA.

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 removed 6 answers which were produced in less than 10 seconds as being unreasonably fast. However, the average time needed to solve our task was 83 seconds. We concluded that the subjects considered our tasks seriously.

6.3. Experimental Results

In both the sandwich game and the multi-attribute road selection game with two-sided uncertainty, we first let the subjects play with the GTBA agent. This agent computes the game theory-based policy of Γ^1, solving the maximization problem presented in section 4. Note that even though the complexity of solving this problem is high, we were able to find the optimal policy for the multi-attribute selection games in a reasonable amount of time.

6.4. Results of the Multi-attribute Road Selection Game with Two-sided Uncertainty

6.4.1. GTBA results. The policy of GTBA using the first settings (Γ^1_ρ) included 13 messages, but 5 of them were generated with a 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.

[4]For a comparison between AMT and other recruitment methods see [Paolacci et al. 2010].

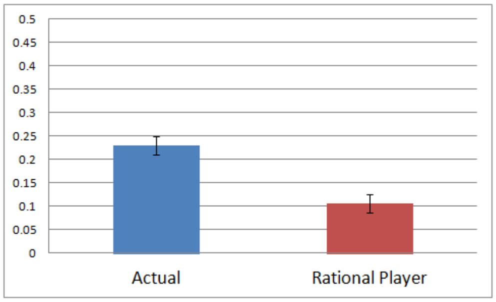

Fig. 1. System utility in road game Γ^1_ρ. The center gained a significantly higher utility from the actual users than the utility it would have gained if all of the users were rational ($p < 0.001$)

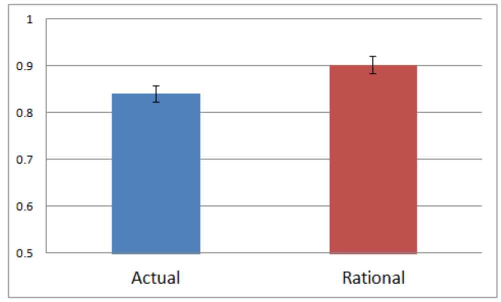

Fig. 2. User utility in road game Γ^1_ρ. The actual drivers gained a significantly lower utility, on average, than they would have gained if they would all act rationally ($p < 0.001$).

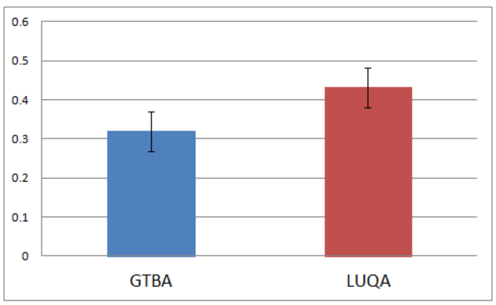

Fig. 3. System utility in road game Γ^2_ρ. The center performed significantly better when using LUQA rather than GTBA ($p < 0.05$).

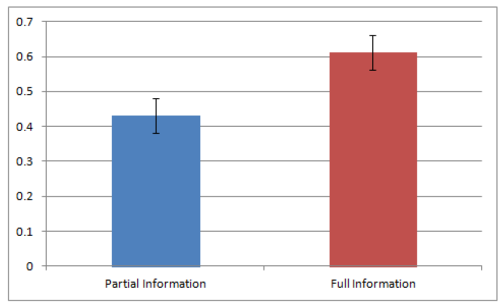

Fig. 4. System utility for LUQA in road game Γ^2_ρ. LUQA performed significantly better when it received full information ($p < 0.05$).

The center received, on average, 0.230 per driver. This result is significantly ($p < 0.001$) higher than the utility that the center would have received if all of the subjects were rational (i.e., maximizing u_r), which, in expectation, was only 0.105 per driver (see Figure 1). As can be seen in Figure 2, user performance significantly dropped from that of fully rational. 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 until 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 -0.015, suggesting that subjects almost ignored the meeting time. These observations lead to the conclusion that in the multi-attribute road selection game with two-sided uncertainty, humans tend to concentrate on the traffic on each road and its toll, but ignore the actual monetary value which supports our general opponent modeling approach for this domain.

6.4.2. LUQ human model. We tested four different methods of modeling human decision-making:

(1) Rational, which assumes that humans always choose the road which maximizes their expected monetary value given in Equation 5. This method is the method assumed by the GTBA agent and does not require any additional parameters.
(2) QRE (logit quantal response), which assumes that the probability that humans choose a road is proportionate to the expected monetary value from that road. This

Table III. Mean square error of modeling human decision-making

Modeling Method	Mean Square Error (the lower the better)
Rational	0.89
QRE	0.295
LWU	0.194
LUQ	**0.065**

method is based on Equation 7, however, it assumes that the drivers base their utility function on the monetary value ($u_r(\vec{v}, a, \theta)$ given in Equation 5) rather than using the subjective utility function ($\bar{u}_\rho(\vec{v}, a)$) (as assumed by LUQ). Therefore, this method has a single parameter: λ.

(3) LWU (Linear Weighted-Utility), which assumes that humans always choose the road which gives them the highest subjective utility (using the subjective utility function $\bar{u}_\rho(\vec{v}, a)$ given in Equation 6. This method has 5 parameters.

(4) LUQ, which combines both linear weighted-utility function and logit quantal response, given in Equations 6 and 7. This method has 5 parameters.

Table III presents the mean square error for all four methods on the data from Γ^1_ρ using a leave-one-out cross validation (in which for each of the messages, when the mean square error is evaluated on a messages, the parameters are learned using data from all other messages). Clearly, LUQ's prediction outperforms all other methods.

6.4.3. Comparing LUQA and GTBA. Using the settings of the second game, Γ^2_ρ, we ran two agents, GTBA and LUQA. We used the results obtained from the 166 subjects that played Γ^1_ρ as the training set data Ψ for LUQA. That is, the αs for $\bar{u}_r^{\Gamma^2_\rho}$ were learned from the subjects playing Γ^1_ρ, i.e., $\mathcal{G} = \{\Gamma^1\}$. LUQA and GTBA each generated 4 messages for Γ^2_ρ. 119 subjects played with LUQA and 102 with GTBA. LUQA performed significantly better ($p < 0.05$) than GTBA, gaining an average of 0.431 vs. 0.319 points per driver (see Figure 3).

We also checked the actual dollars earned by the subjects. Unfortunately, when playing with LUQA the average virtual gain per subject was only $\$19.00$, while when playing with GTBA the average was higher, $\$21.20$. These results differ significantly, hinting that the center's gain was on account of the driver's monetary utility. This result is compatible with our previous result in [Azaria et al. 2012a], where people tend to perform better when the agent confronting them assumes that they will act rationally. However, in practice, this issue isn't of great concern, since, if the center is interested in the driver receiving a higher utility, it may implicitly add the driver's utility to its own utility function and result with a protocol that will be better for both the center and the driver.

6.4.4. One-sided uncertainty vs. Two-sided uncertainty. In previous work [Azaria et al. 2011] we tested the performance of LUQA in the road selection problem under the exact same settings, only with full information for the center. Figure 4 shows these results along with our current results with partial information. As can be seen, when LUQA has full information, it significantly outperforms LUQA with partial information. This is not surprising, since additional information allows the Sender to avoid mistakes and encourages the Receiver to take actions which are more favorable to the Sender.

6.5. Sandwich Game Results

6.5.1. GTBA results. The monetary result plays an important role in the sandwich game. This is because the game is played in an environment a person's goal is to make

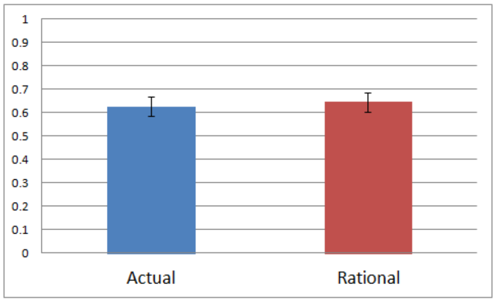

Fig. 5. User utility in sandwich games. The difference between a fully rational seller and the actual human sellers is minor and not statistically significant.

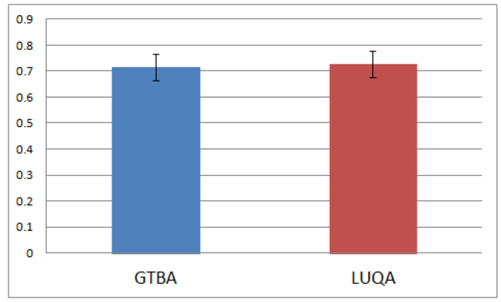

Fig. 6. System utility in sandwich game Γ_σ^2. The difference between the organizer utility when using LUQA and when using GTBA is minor and not statistically significant.

as high a revenue as possible, which usually results in selling as many sandwiches as possible while minimizing the number of sandwiches thrown away.

The policy of GTBA in the first settings (Γ_σ^1) included 5 messages. The organizer received on average 0.260 per seller (Figure 7). The utility of the organizer was similar to the expected utility that the organizer would receive if all subjects were rational (i.e., maximizing u_r), which, in expectation, was 0.299 per seller. We suspect that this is due to the important role that the monetary value plays in this game. These results differ from the correspondence results of the road selection game and thus we hypothesis that LUQA is not needed here and that the GTBA agent will do as well as LUQA in this domain.

6.5.2. LUQA and the LUQ human model. The learning phase for LUQA, which was based on the subjects which participated in Γ_σ^1, found the following parameters in the subjective utility function: α_1, which is the amount gained by each sandwich sold, is 0.087, and α_2, which is the amount lost by each sandwich thrown away was -0.103. On average (depending on the private type w), if maximizing expected monetary values, people should have been neutral between missing a sandwich and preparing one too many sandwiches, however, it seems that people were a little risk averse since $|-0.103| > |0.087|$, but still the numbers are very close. When testing the MSE of LUQ, we get a result similar to that of QRE (quantal response under expected monetary outcome), both yielding 0.07. The similar performance for both LUQ and MSE indicates that people performed nearly rationally, which nearly obviates the usage of LUQ.

6.5.3. Comparing LUQA and GTBA. The comparison was done under the second settings (Γ_σ^2), and both GTBA and LUQA used 4 messages. The organizer received on average 0.715 when using GTBA, and 0.728 when using LUQA (Figure 6). Although LUQA did perform slightly better, these results do not differ significantly. This is not surprising since, as mentioned above, the subjects' subjective utility was very close to the expected monetary value and thus the GTBA's assumptions were correct. We suggest that the slight improvement shown was given from the logit quantal response assumption.

6.6. Deciding between LUQA and GTBA

As demonstrated in the above two games, there are situations where LUQA outperforms GTBA, but in some situations they yield similar results. One may recommend to always use LUQA since it is always as good as GTBA and sometimes even better. However, LUQA requires collecting data to learn the human utility function. Therefore, we recommend to first collect some data using GTBA and compare the agent's results and

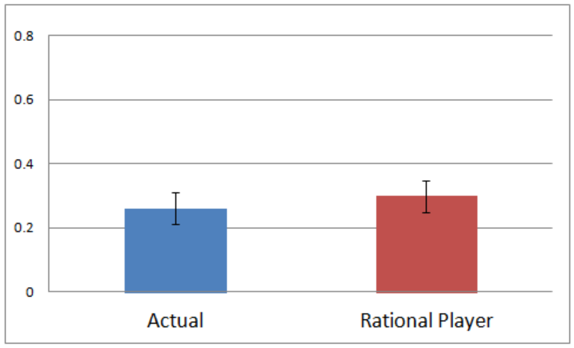

Fig. 7. System utility in sandwich game. The difference in the organizer's utility between actual users and the utility it would have gained if all of the users were rational is minor and not statistically significant.

the human behavior to the rational behavior. If the GTBA's results are significantly different from the expected results that it would have received if people would have followed a rational decision-making process, then it is worthwhile to collect more data and use LUQA. Otherwise, using GTBA seems to be a good enough heuristic. It is important to note that in the road selection game it was enough to use 10 subjects in order to obtain a significant difference between GTBA and the expected utility if people would have followed rational behavior. In the sandwich game, we didn't obtain significant results even with 100 subjects. So collecting $20-25$ data points for making a decision whether to use LUQA or GTBA seems reasonable.

7. CONCLUSIONS

In this paper we consider information disclosure games with two-sided uncertainty 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 with two-sided uncertainty, assuming that the human is fully rational. We also provide an innovative machine learning-based model that effectively predicts people's behavior in these games. The model we provide assumes that people use a subjective utility function which is a linear combination for all given attributes. The model also assumes that while people use this function as a guideline, they do not always choose the action with the greatest utility value, however, the higher an action's utility value is, the more likely they are to choose that action. We integrate this model into our persuasion model in order to yield an innovative method of human behavior manipulation. Extensive empirical study in multi-attribute road selection games with two-sided uncertainty confirms the advantage of the proposed model in that game. However, in another domain we tested, the Sandwich game, there is no significant advantage to the machine learning-based model, and using the game theory-based agent which assumes that people maximize their expected monetary values is beneficial. We propose a methodology of how to choose between the two options. We argue that, depending on the domain, people's decision-making process may vary and thus where in one domain modeling humans as rational may be good enough, in another domain this model is too far from their actual behavior and therefore an agent that assumes perfectly rational behavior may fall far behind.

REFERENCES

Amazon. 2013. Mechanical Turk Services. http://www.mturk.com/. (2013).

Amos Azaria, Yonatan Aumann, and Sarit Kraus. 2012. Automated Strategies for Determining Rewards for Human Work.. In *AAAI*.

A. Azaria, Z. Rabinovich, S. Kraus, and C. V. Goldman. 2011. Strategic Information Disclosure to People with Multiple Alternatives. *Proc. of AAAI* (2011).

A. Azaria, Z. Rabinovich, S. Kraus, C. V. Goldman, and Y. Gal. 2012a. Strategic Advice Provision in Repeated Human-Agent Interactions. *AAAI* (2012).

A. Azaria, Z. Rabinovich, S. Kraus, C. V. Goldman, and O. Tsimhoni. 2012b. Giving Advice to People in Path Selection Problems. *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (2012), 459–466.

S. Bonaccio and R. S. Dalal. 2006. Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes* 101, 2 (2006), 127–151.

C. F. Camerer. 2003. *Behavioral Game Theory. Experiments in Strategic Interaction*. Princeton University Press, Chapter 2.

C. G. Chorus, E. J. Molin, and B. van Wee. 2006. Travel information as an instrument to change car drivers travel choices: a literature review. *EJTIR* 6, 4 (2006), 335–364.

V. Crawford and J. Sobel. 1982. Strategic Information Transmission. *Econometrica* 50 (1982), 1431–1451.

N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, and R. J. Dolan. 2006. Cortical substrates for exploratory decisions in humans. *Nature* 441, 15 (2006), 876–879.

M. Fenster, I. Zuckerman, and S. Kraus. 2012. Guiding User Choice During Discussion by Silence, Examples and Justifications. *Proc. of ECAI* (2012).

Y. Gal and A. Pfeffer. 2007. Modeling reciprocal behavior in human bilateral negotiation. *Proceedings of the National Conference on Artificial Intelligence* 22, 1 (2007).

J. Glazer and A. Rubinstein. 2006. A Study in the Pragmatics of Persuasion: A Game Theoretical Approach. *Theoretical Economics* 1 (2006), 395–410.

Google. 2013a. Google Maps. http://maps.google.com/. (2013).

Google. 2013b. WAZE. http://www.waze.com/. (2013).

Chen Hajaj, Noam Hazon, David Sarne, and Avshalom Elmalech. 2013. Search More, Disclose Less. In *Proc. of AAAI*.

R. A. Horn and C. R. Johnson. 1991. *Topics in Matrix Analysis*. Cambridge University Press.

P. Hoz-Weiss, S. Kraus, J. Wilkenfeld, D. R. Andersend, and A. Pate. 2008. Resolving crises through automated bilateral negotiations. *Artificial Intelligence Journal* 172, 1 (2008), 1–18.

E. Kamenica and M. Gentzkow. 2010. *Bayesian persuasion*. Technical Report. University of Chicago. under review.

X. J. Kuang, R. A. Weber, and J. Dana. 2007. How effective is advice from interested parties? *J. of Economic Behavior and Organization* 62, 4 (2007), 591–604.

D. Lee. 2006. Best to go with what you know? *Nature* 441, 15 (2006), 822–823.

W. B. G. Liebrand. 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 (1984), 239–264.

H. S. Mahmassani and Y.-H. Liu. 1999. Dynamics of commuting decision behaviour under advanced traveller information systems. *Transporation Research Part C: Emerging Technologies* 7, 2-3 (1999), 91–107.

P. Milgrom and J. Roberts. 1986. Relying on the Information of Interested Parties. *Rand J. of Economics* 17 (1986), 18–32.

T. Nguyen, R. Yang, A. Azaria, S. Kraus, and M. Tambe. 2013. Analyzing the effectiveness of adversary modeling in security games. In *Conf. on Artificial Intelligence (AAAI)*.

M. J. Osborne and A. Rubinstein. 1994. *A course in Game Theory*. MIT Press.

Y. Oshrat, R. Lin, and S. Kraus. 2009. Facing the Challenge of Human-Agent Negotiations via Effective General Opponent Modeling. In *Proc. of AAMAS 2009*.

G. Paolacci, J. Chandler, and P. G. Ipeirotis. 2010. Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making* 5, 5 (2010).

N. Peled, Y. Gal, and S. Kraus. 2011. A Study of Computational and Human Strategies in Revelation Games. (2011).

E. Turkay Pillai. 2009. Two-Sided Uncertainty in Persuasion Games, working paper. (2009).

Luis Rayo and Ilya Segal. 2010. Optimal Information Disclosure. *Journal of Political Economy* 118, 5 (2010), 949–987.

A. Rosenfeld and S. Kraus. 2011. Using Aspiration Adaptation Theory to Improve Learning. (2011).

Avi Rosenfeld, Inon Zuckerman, Amos Azaria, and Sarit Kraus. 2012. Combining psychological models with machine learning to better predict peoples decisions. *Synthese* 189, 1 (2012), 81–93.

David Sarne, Avshalom Elmalech, Barbara J Grosz, and Moti Geva. 2011. Less is more: restructuring decisions to improve agent search. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 431–438.

I. Yaniv and E. Kleinberger. 2000. Advice Taking in Decision Making: Egocentric Discounting and Reputation Formation. *Organizational Behavior and Human Decision Processes* 83, 2 (November 2000), 260–281.

A. PROOFS OF THEOREMS REGARDING THE MESSAGE SPACE

A.1. Proof of Theorem 4.1

PROOF.

Let (π, M) be an optimal solution to the game so that $|M| > |\Omega| = n$. We will first show that certain transformations of π produce a left stochastic matrix structure (in which the rows correspond to messages and the columns to observations) with at least one zero row, i.e. produce disclosure rules that use less messages than the original π. We will then show a specific transformation of π that, while reducing the number of used messages, preserves the utility gained. We will thus obtain a new optimal disclosure rule with fewer messages. Since $|M| < \infty$, iterative application of the above process would lead to an optimal $(\widetilde{\pi}, \widetilde{M})$, where $|\widetilde{M}| <= n$ as required.

Notice again that zero rows in π correspond to the messages that are never sent, and we would be able to reduce the size of M without changing the utility in any way. Assume that after the elimination of zero rows, we still have the set of messages greater than $|\Omega|$ or there never were any.

Since π is a stochastic matrix, there can be no more than n elements in it equal to 1. If all are present, the rest of the rows are zero, and we can reduce M to have only n elements without changing π, hence obtaining the necessary optimal solution properties. If this does not occur, i.e. there are less than n elements in π equal to 1, we can proceed with the following reasoning.

Denote π_m the m'th row of π. It holds $\sum\limits_{m \in M} \pi_m = \vec{1}_n^T$, where $\vec{1}_n$ is a column vector in R^n with all elements equal to 1. Since there are at least $n+1$ rows in π, but only n columns, π has a non-trivial kernel space of left multiplication vectors. Hence, there is a non-trivial row vector $\alpha = (\alpha_m)_{m \in M}$ so that $\phi = \alpha\pi = \vec{0}_n^T$, and for all $m \in M$ $|\alpha_m| \leq 1$ and for some $m_1 \in M$ $\alpha_{m_1} = 1$. This can be achieved by taking an arbitrary non-trivial kernel row vector and scaling it appropriately.

Clearly, $\pi_{m_1}(\omega) < 1$ for all $\omega \in \Omega$. Otherwise, for some $\bar{\omega} \in V$ $\pi_m(\bar{\omega}) = 0$ for all $m \neq m_1$, and $\phi(\bar{\omega}) = \alpha_{m_1} = 1 \neq 0$, hence contradicting $\phi = \vec{0}$.

Denote $\widetilde{\pi}$ a matrix with rows defined by $\widetilde{\pi}_m = (1 - \alpha_m)\pi_m$. Notice that all elements of $\widetilde{\pi}$ are non-negative. Furthermore, they are not greater than 1, due to the following:

$$\begin{aligned}
\vec{1}_n^T &= \sum_{m \in M} \pi_m = \vec{1}_{|M|}^T \pi \\
\vec{0}_n^T &= \alpha\pi \\
\vec{1}_n^T &= (\vec{1}_{|M|}^T - \alpha)\pi = \vec{1}_{|M|}^T \widetilde{\pi}
\end{aligned}$$

Since all elements of $(\vec{1}_{|M|}^T - \alpha)$ are non-negative, and so are elements of π, the last equation means that elements of $\widetilde{\pi}$ are bounded by 1, and the sum of rows is $\vec{1}_n^T$.

Hence $\widetilde{\pi}$ is also a valid solution to the game. Furthermore, it uses less messages since $\alpha_{m_1} = 1$ and $\widetilde{\pi}_{m_1} = (1 - \alpha_{m_1})\pi_{m_1} = \vec{0}_{|M|}$.

Applying the above reasoning in an iterative fashion, we can reduce the number of non-zero rows in $\widetilde{\pi}$ to n. Denote $\widetilde{M}$ to be the subset of M that corresponds to those rows, this will be the new set of messages.

We will now show that $\widetilde{\pi}$ has the same utility as π, hence $(\widetilde{\pi}, \widetilde{M})$ will be an optimal disclosure rule with $|\widetilde{M}| = n$, concluding the proof.

Denote $U_s^M[\pi_m] = \sum_{v \in V} \sum_{a \in A} \sum_{\omega \in \Omega} u_s(v,a) p_V(v) p_A(a|p_V^m) p_\Omega(\omega|v) \pi(m|\omega)$, then $U_s[\pi] = \sum_{m \in M} U_s^M[\pi_m]$. Notice that $U_s^M[\gamma \pi_m] = \gamma U_s^M[\pi_m]$, since p_V^m is insensitive to scaling of π_m.

Let us now compute $U_s[\widetilde{\pi}]$, where $\widetilde{\pi}$ was computed using a vector $\gamma\alpha$ with $\gamma \in R$.

$$
\begin{aligned}
U_s[\widetilde{\pi}] &= \sum_{m \in M} U_s^M[\widetilde{\pi}_m] \\
&= \sum_{m \in M} U_s^M[\widetilde{\pi}_m] \\
&= \sum_{m \in M} U_s^M[(1 - \gamma\alpha_m)\pi_m] \\
&= \sum_{m \in M} (1 - \gamma\alpha_m) U_s^M[\pi_m] \\
&= U_s[\pi] - \gamma \sum_{m \in M} \alpha_m U_s^M[\pi_m] \\
&= U_s[\pi] - \gamma * U_{diff}
\end{aligned}
$$

If $U_{diff} \neq 0$, then for $\widetilde{\pi}$ computed for $\gamma = \mathbf{sign}(U_{diff})$ we have that $U_s[\widetilde{\pi}] \gneq U_s[\pi]$, hence contradicting the optimality of π. Therefore, $U_{diff} = 0$ and (setting $\gamma = 1$) we have $U_s[\widetilde{\pi}] = U_s[\pi]$, making $(\widetilde{\pi}, \widetilde{M})$ an alternative optimal solution with $|\widetilde{M}| = |\Omega|$, as required.

□

A.2. Proof of Theorem 4.2

PROOF. The following proof is stated for a countable infinity of messages. However, since the space of all possible conditional message probabilities π_m is compact, it is easy to recast it for continuous message indices.

Let (π, M) be an optimal solution to the problem, so that $|M| = \infty$, and furthermore for an infinite number of messages $\pi_m p_V > 0$. In other words there is an infinite number of messages that have a non-zero probability to appear, w.l.g. assume that all messages are such. Notice also that w.l.g. we can assume that $u_s(v,a) > 0$ for all $a \in A$ and $v \in V$. Denote $u_s^{min} = \inf_{p(v,a) \in \Delta(V \times A)} \mathbf{E}[u_s(v,a)] > 0$, and notice that $U_s^M[\pi_m] \geq u_s^{min}$ for any π_m. Similarly notice that $u_s^{max} = \sup_{p(v,a) \in \Delta(V \times A)} \mathbf{E}[u_s(v,a)] < \infty$ and that $U_s^M[\pi_m] \leq u_s^{max}$.

Since the sum of all message probabilities is equal to 1, and all utilities are strictly positive, the sequence of partial sums $\sum_{i=0}^{t} U_s^M[\pi_{m_j}]$ is monotonic increasing and bounded, hence $U_s[\pi] = \sum_{i=0}^{\infty} U_s^M[\pi_{m_i}] < \infty$ and is well defined. Furthermore, for any $\epsilon > 0$ exists $T < \infty$ so that $\sum_{i=T+1}^{\infty} U_s^M[\pi_{m_i}] \leq \epsilon$. Consider setting $\epsilon = u_s^{min}$ and set $\widetilde{\pi}_m = \begin{cases} \pi_{m_i} & m = m_i, i \in [0:T] \\ \sum_{i=T+1}^{\infty} \pi_{m_i} & m = \widetilde{m} \end{cases}$. It holds that $U_s^M[\widetilde{\pi}_{\widetilde{m}}] \geq u_s^{min} \geq \sum_{i=T+1}^{\infty} U_s^M[\pi_{m_i}]$. Therefore, $U_s[\pi] \leq U_s[\widetilde{\pi}]$, and $(\widetilde{\pi}, \widetilde{M})$ is a finite disclosure rule with a utility at least as good as the original solution (π, M). Hence, if the optimal U_s is obtainable, then there is a finite disclosure rule that achieves it. $\square$