# Advice Provision for Choice Selection Processes with Ranked Options

**Amos Azaria**[1] and **Ya'akov Gal**[2] and **Claudia V. Goldman**[3] and **Sarit Kraus**[1,4]
[1] Department of Computer Science, Bar-Ilan University, Ramat Gan 52900, Israel
Amos Azaria Website: http://azariaa.com
[2] Department of Information Systems Engineering, Ben-Gurion University of the Negev, Israel
[3] General Motors Advanced Technical Center, Herzliya 46725, Israel
[4] Institute for Advanced Computer Studies University of Maryland, MD 20742

## Abstract

Choice selection processes are a family of bilateral games of incomplete information in which a computer agent generates advice for a human user while considering the effect of the advice on the user's behavior in future interactions. The human and the agent may share certain goals, but are essentially self-interested. This paper extends selection processes to settings in which the actions available to the human are ordered and thus the user may be influenced by the advice even though he doesn't necessarily follow it exactly. In this work we also consider the case in which the user obtains some observation on the sate of the world. We propose several approaches to model human decision making in such settings. We incorporate these models into two optimization techniques for the agent advice provision strategy. In the first one the agent used a social utility approach which considered the benefits and costs for both agent and person when making suggestions. In the second approach we simplified the human model in order to allow modeling and solving the agent strategy as an MDP. In an empirical evaluation involving human users on AMT, we showed that the social utility approach significantly outperformed the MDP approach.

## Introduction

Many real life situations require computer agents to generate advice to human users concerning which actions to perform. In most cases, it is assumed that the computer agent possess more information than the user. Not always do the computer agent and human user share the exact same goal. One example for such a situation, is a computer agent that advises a user which route to commute. While the user is more interested in minimizing his travel time, the system may be more interested in minimizing the user's air pollution. Another example may be a system that advises a user on the temperature to set on a electric car's Air Conditioner (AC) system on a sunny day. While the user may want the AC to be as strong as possible and not care as much about the battery consumption, the system may care more about the latter.

In previous work (Azaria et al. 2012), we introduced a family of repeated games of incomplete information called *choice selection processes* that include two players: an agent "Sender" and a human "Receiver". The Sender, which possess more information than the Receiver, needs to advise the Receiver which action to perform. The players are essentially self-interested, but their goals may interact. We considered problems motivated by repeated route selection settings that can be modeled as repeated multi-arm Bandit problems and showed that an agent that combined a hyperbolic discounting model of human behavior with a social utility function (SAP) was able to outperform alternative agent designs when empirically evaluated with people. In particular, modeling the problem as a continuous MDP and solving it using Markov Chain Monte Carlo sampling methods yielded lower utility to the agent than SAP.

In this work, we study a different type of choice selection processes which is motivated by the AC temperature selection problem which are distinct in the following ways:

- Ordered actions: The action set is an ordinal scale which represents the energy consumption level of the CCS in the car. The roads in the previous domain that we examined were not sorted in any scale. The actions were a set of not-ordered options.
- Partial acceptance: The receiver may *partially* accept the advice (e.g. set the power level of the AC to a lower level than initially intended, but not as low as suggested by the sender). This makes the task of modeling the receiver significantly more difficult. In the roads domain, a user could either accept or not the advice; in the climate control case a user can partially accept an advice and in a sense make a choice that takes him closer to the advice.
- Cost for Receiver: The cost to the receiver depends on two attributes: the energy consumption of the CCS, and the user's comfort level which depends on the energy consumption and the state of the world. Therefore, modeling the human behavior becomes a more complex task than the task of modeling humans in the roads domain.
- Partial observability: The receiver is given an observation (the heat load) that is associated with the state of the world. Therefore it is able to update its belief over the state of the world. In the roads domain, the user was assumed not to have any information about the traffic distribution in the different roads.
- Finite state space: The configuration constrains the state space to solve the MDP. In the roads domain, the state space was larger and it was not practical to find the optimal solution to the corresponding MDP.

We study this problem in the context of a game in which modeling the Receiver's decision-making becomes a harder problem but the Sender's uncertainty about future state of the world is drastically reduced, and thus building an MDP to solve the Sender's problem becomes feasible. Never the less, we show that SAP strategy yields significantly lower cost to the Sender than the MDP strategy.

## The AC Game

We study a choice selection processes which differs from the route selection domain described in (Azaria et al. 2012) and includes an imaginary scenario where a car driver needs to set how much power he would like his Climate Control System (CCS) to consume. We denote this by "power level" of the CCS.

Higher values of the power level are associated with increased energy consumption by the system. The sender player represents a system which suggests a power level setting to the receiver (the driver). At each round, the receiver is given a discrete observation $o(v)$ that represents the current heat load in the car (a function of temperature, humidity and other environmental conditions).

## Experiments

All of our experiments were performed using Amazon's Mechanical Turk service (AMT). Participation in all experiments consisted of a total of 272 subjects from the USA, of which $44.4\%$ were females and $55.6\%$ were males. The subjects' ages ranged from 19 to 67, with a mean of 32.3 and median of 30. All subjects had to pass a short quiz to assure that they understood the game.

The first set of experiments were designated to collect data for modeling human behaviour in the game. We randomly divided subjects into one of several treatment groups. The user model that best fitted the data was composed of four elements: (i) to what extent the user relies on the advice and is based on hyperbolic discounting (similar to the approach in (Azaria et al. 2012). (ii) the impact of the advice along with the user's reliance on it (obtained from the first element). The impact of the advice is assumed to not only be on the action advised, but also on close actions (an exponential decay). (ii) the learning curve of which the users learned their actual true score. (iv) the actual weights that humans apply to the AC energy consumption level and the comfort level which differ from the user's actual score (A similar approach was successfully used in (Azaria et al. 2011)). As in (Azaria et al. 2012), the user is assumed to use logit quanatal response when selecting his action.

Based on this user model we constructed two agents. In the SAP agent the human model was used for simulating the Receiver's decision making in order to determine the weights of the subjective utility function, i.e. the coefficients of the linear combination on both the user and the agent's cost functions, which yielded the lowest total cost for the agent. In the MDP-based agent we simplified the Receiver's model by using exponential smoothing rather than Hyperbolic discounting and replaced the "learning"' component by the assumption that people know their exact comfort level (but still use their own weights when considering an action). This simplification only slightly decreased the fitness of the model to the collected data but allowed us to solve the MDP (using value-iteration). As base-line we also tested subjects three treatments: Subjects in the *Silent* group received no advice at all. Subjects in the *Receiver* group were consistently advised to choose the AC energy consumption level that was most beneficial to them, (i.e., associated with the highest score). Lastly, subjects in the *Sender* group were consistently advised to choose the AC level which was best for the Sender (i.e., were constantly advised to set the AC energy consumption level to 1). Table 1 presents the average performance for each of the groups. SAP significantly ($p < 0.001$ using ANOVA test) outperformed all other methods (including the MDP method). The MDP method also performed quite well, as it outperformed all other baselines. Unsurprisingly, the subjects in the *Receiver* group yield the best score, however, the acceptance rate of both the MDP method and SAP were very close to that of the subjects following the advice in the *Receiver* group.

| method | energy consumption | comfort level | user score | acceptance |
|---|---|---|---|---|
| Silent | 5.202 | 8.744 | 12.289 | – |
| Receiver | 5.197 | 8.933 | **12.67** | **36.7%** |
| Sender | 4.437 | 7.843 | 11.264 | 19.5% |
| SAP | **3.952** | 7.466 | 11.02 | 34.5% |
| MDP | 4.361 | 7.996 | 11.652 | 33.8% |

Table 1: Performance Results against People

## Discussion and Conclusions

In this paper we introduce the AC game and describe a method of modeling human decision making process in such a complex domain. We assimilate this model into SAP, an agent presented in (Azaria et al. 2012), in order to provide advice to the user. Furthermore, the nature of the AC game, was designed in a manner that allowed construction of an MDP. While, the MDP method, did outperform other baseline methods, still, SAP outperformed all methods including the MDP. It may seem surprising that SAP, which uses a relatively simple method outperformed the MDP approach. We explain this by the fact that the user model had to be simplified and discretized in order to suite the MDP. Furthermore, a human model may never be exact, therefore, relying too much on a noisy model as done by the MDP, may not perform as well as SAP, which only uses the human model as a guideline.

## References

Azaria, A.; Rabinovich, Z.; Kraus, S.; and Goldman, C. V. 2011. Strategic information disclosure to people with multiple alternatives. In *AAAI*.

Azaria, A.; Rabinovich, Z.; Kraus, S.; Goldman, C. V.; and Gal, Y. 2012. Strategic advice provision in repeated human-agent interactions. In *AAAI*.