Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)

Intelligent Advice Provisioning for Repeated Interaction

Priel Levy
Bar-Ilan University, Israel
levypri1@cs.biu.ac.il

David Sarne
Bar-Ilan University, Israel
sarned@cs.biu.ac.il

Abstract

This paper studies two suboptimal advice provisioning methods ("advisors") as an alternative to providing optimal advice in repeated advising settings. Providing users with suboptimal advice has been reported to be highly advantageous whenever the optimal advice is non-intuitive, hence might not be accepted by the user. Alas, prior methods that rely on suboptimal advice generation were designed primarily for a single-shot advice provisioning setting, hence their performance in repeated settings is questionable. Our methods, on the other hand, are tailored to the repeated interaction case. Comprehensive evaluation of the proposed methods, involving hundreds of human participants, reveals that both methods meet their primary design goal (either an increased user profit or an increased user satisfaction from the advisor), while performing at least as good with the alternative goal, compared to having people perform with: (a) no advisor at all; (b) an advisor providing the theoretic-optimal advice; and (c) an effective suboptimal-advice-based advisor designed for the non-repeated variant of our experimental framework.

Introduction

One important role of collaborative interfaces and AI-based systems is supporting people in decision situations by providing them beneficial advices or suggesting a preferred course of action (Azaria, Kraus, and Richardson 2013; Rosenfeld et al. 2015; Ricci et al. 2011). Most research to date in the area of advice provisioning focused on providing the optimal (e.g., benefit-maximizing or utility-maximizing) advice, making the assumption that the advice provided would definitely be accepted by the user (Bharati and Chaudhury 2004). Yet, there is much evidence that people often fail to see the benefit in and often ignore utility-maximizing advices and hints, especially when those are associated with non-intuitive choices or courses of action (Carroll, Bazerman, and Maury 1988). Therefore, an effective advice is not necessarily the one encapsulating the greatest benefit to the user if accepted, but rather one that maximizes the expected benefit when taking into consideration its likelihood to be adopted by the user and its effect over the user's choice otherwise (Elmalech et al. 2015; Azaria et al. 2014).

Alas to date, most designs following the above paradigm have been focused in one-shot decision situations. As such they are completely stateless and static in the sense that they always provide the same advice at a given decision situation. Nevertheless, for many collaborative agents supporting people through advice provisioning the interaction with the user is inherently repeated, and consequently their effectiveness is measured over the long term. Examples include route planning (e.g., googlemaps.com), investments advising (e.g., betterment.com) and recommending the timing for buying a flight ticket (e.g., in kayak.com). As discussed in detail in the following section, the repeated interaction dictates designs that consider alongside the likelihood that the user will accept a given advice also the influence of that advice (and the possible different outcomes if followed) over the user's willingness to accept future advices from the advisor. For that reason, methods designed for providing suboptimal advice in non-repeated settings perform poorly once switching to repeated settings—in our experiments the performance of such method was even worse than the performance of an advisor providing the theoretic-optimal advice, reversing the dominance relationship between the two that was demonstrated in the non-repeated version of the setting used (Elmalech et al. 2015).

This paper studies two new methods for advice-provisioning exclusively aimed at repeated interaction settings. Both methods are designed to implicitly improve the value the user sees in the advices provided to her over time. Yet, despite this similar underlying design principle, each method attempts to maximize a different measure. The first method is aimed at maximizing the user's actual *profit* when provided with the advisor, through a gradual convergence to the optimal (non-intuitive) advice. The second method aims at maximizing the *user's satisfaction* from the advisor through considering a regret-like index. As our experimental results show, there is only a small correlation between profit maximization and user satisfaction in repeated advice settings, hence the importance of having the two methods.

The methods were tested using extensive experimentation with 500 human participants overall, using a well-established testbed that enabled a common ground for comparing the performance of our methods with others and for partial validation of the experimental design, as discussed in detail in later sections. The analysis of the results obtained indicates that both proposed methods meet their design goals, exhibiting a statistically significant improved performance according to the measure they are focused in (user profit or user satisfaction) compared to all other tested methods (performing with no advisor, with an advisor always providing the optimal advice and with an efficient (experimentally-provable) advisor for the non-repeated setting) while keeping at least the same level of performance according to the alternative measure.

In the following section we describe in detail the idea of suboptimal advising in the context of effective advice provisioning. Then, we describe our proposed methods for suboptimal advising in repeated-interaction settings. The descriptions of the experimental infrastructure, the experimental design and the results obtained follow. Finally we conclude with a discussion of the findings and directions for future research. Related work is cited throughout the paper.

Effective Advice Provisioning

Assume an advisor is capable of extracting the strategy that is optimal to the user (taking into consideration all possible outcomes and transitions to different world states), denoted "optimal strategy" onwards.[1] Sticking to providing advice according to the optimal strategy is best only if the advisor is fully confident that its advice is always adopted by the advisee. Alas, people often do not follow the theoretic-optimal strategy (Zhang, Bellamy, and Kellogg 2015; Hajaj, Hazon, and Sarne 2015), especially in domains where the optimal strategy has some non-intuitive properties (Rochlin, Sarne, and Mash 2014; Rochlin and Sarne 2014), nor use it when designing agent strategies (Rosenfeld and Kraus 2012; Rosenfeld et al. 2012). This phenomenon carries over to receiving advice and it has been shown that providing the optimal advice in such domains typically results in low acceptance rate and consequently poor outcomes (Elmalech et al. 2015; Rosenfeld and Kraus 2015). Various approaches were taken in prior work in order to overcome people's inability to recognize the benefit in following the optimal (or socially-optimal) strategy, such as attempting to convince people of the correctness of the optimal solution (Grosskopf, Bereby-Meyer, and Bazerman 2007) or to teach people how to optimally solve the decision problem (Lesser 1999). These approaches are irrelevant for our repeated advice provisioning settings because they require substantial overhead from the advisor's side (and much of the user's time).

Our approach coincides with the recently introduced idea of using designs that provide suboptimal advices, yet ones that are more appealing to the user thus are more likely to be adopted (Elmalech et al. 2015). Previous work that has used the idea aimed to non-repeated settings.[2] As such, the main aspects considered in the designs used were the questions of when agents should deviate from dispensing fully rational advice and how far should the suboptimal advice provided drift from the optimal one. In repeated settings the user gets to see, and naturally becomes influenced by, outcomes of (and regret from) previous advices provided by the advisor over time, making the advisor's design more challenging.[3] Therefore the design of an advisor for such environments should also consider aspects such as the level by which any specific advice provided at a given point will influence the user's tendency to adopt further advice provided by the advisor. In particular, it should balance to some extent between the efforts to push the user to accept a given advice (or act in a certain prescribed way) and the relative loss due to compromising on suboptimal advice. For example, following a temporal sequence of losses in a poker game the advisor may need to adopt a more conservative betting strategy, even though it is not the expected-profit-maximizing strategy at that time, in order to regain the user's confidence and increase the chance of consecutive advices to be accepted.

Notice that up to this point we have been using the term "optimal" without discussing the essence of optimality. There are several candidates for measuring the optimality of a decision, ranging from utility theory and its many variants (Rochlin and Sarne 2013) to theories that incorporate decision weights that reflect the impact of events on the overall attractiveness of uncertain choices (Starmer 2000), such as sign-dependent theories (e.g., prospect theory (Kahneman and Tversky 1979)) and rank-dependent theories (Quiggin 1982). Common to all these theories is that they require a deep understanding of individual preferences and possibly comprehensive utility elicitation efforts if basing the optimality measurement upon them. Fortunately, when it comes to repeated interaction the optimality measure is simpler to extract, as people's tendency to rely on expectations in this case substantially strengthen. It has been shown that in repeated-play settings people's strategies asymptotically approach the expected monetary value (EMV) strategy as the number of repeated plays increases (Klos, Weber, and Weber 2005). Within this context, prior work attributes people's failure to use an EMV-maximizing strategy to the complexity and unintuitive nature of this strategy (Elmalech, Sarne, and Grosz 2015). We thus take these evidence as a basis for assuming that people are EMV-maximization seeking in repeated decision situations and use this strategy as the measure of optimality in analyzing our experimental results.

Relying on EMV as the sole measure of performance for our methods is of course incomplete, as even if an advisor manages to maximize (or at least substantially improve) the expected payoff with the advices it provides, there is no guarantee that the user perceives it as such. Since nature plays a role in our settings, it is possible that even expected-payoff-maximizing strategy will result in a sequence of non-favorable actual outcomes at times. Users, as discussed above, are highly affected by the nature of outcomes (e.g.,

[1]For now, we intentionally defer the discussion concerning the nature of "optimality".

[2]An exception is the work of Azaria et al. (2015) however the interaction setting there is very different from ours, as their use of a non-optimal advice is primarily to push the user towards a choice that is more beneficial for the advisor itself. Also, their advice generation relies on the asymmetry in the information available to the advisor and the user, while in our setting the asymmetry is in the ability to identify the optimal advice.

[3]Feedback plays a key role in decision making, as demonstrated in prior work (Jessup, Bishara, and Busemeyer 2008).

gains versus losses) and effects such as regret, and thus may exhibit different levels of satisfaction based on the advices given and actual outcomes, such that are not fully correlated with the EMV. Therefore, to provide a comprehensive evaluation of user satisfaction we use two complementary measures. The first is the subjective reporting of the users themselves regarding their overall satisfaction from the advisor provided to them. Decision maker's satisfaction has traditionally functioned as a measure of past and current utility (Lemon, Barnett White, and Winer 2002) and hence its key influence on keeping or dropping the advice given. The second is the users' adoption rate of the advices provided by the advisor, measured as the ratio between the number of advices adopted and the overall number of advices provided. This latter measure is important for two reasons. First, unlike the first measure it is an objective measure. Second, it is a measure of success for many real-world advisors, whenever the advisor is compensated based on the number of times its advice is accepted or used, typically in the form of commission (e.g., financial advisors and investments brokers).

Proposed Advice Generation Methods

Both our advice-generation methods are based on providing suboptimal advice, however differ in the emphasis they place on the different aspects of the advising effectiveness, i.e., actual performance and user satisfaction.

S-Gradual With this method the advisor always starts with what might seem to be a highly intuitive advice for people in the specific decision situation. A "highly intuitive advice" can be generated either based on observing decisions made by other people when facing a similar decision situation (Martin et al. 2004) or by constructing user models that take into consideration various kinds of known psychological effects and human behaviors (Hajaj, Hazon, and Sarne 2014) to predict the appeal of different advices to people in general. The idea is to strengthen the user's perception that the advisor "knows what it is doing" and then gradually trade intuitiveness with expected benefit. Meaning that on each interaction the generated advice becomes closer to the optimal one, encompassing greater expected profit to the user (if adopted), though with the price of an increased non-intuitiveness. The underlying assumption here is that over time the user will become less resistant to "unsensible" advices due to prior successful "more sensible" ones. Despite the simplicity of the idea, to the best of our knowledge it has not been used in the design of advice-provisioning systems.[4]

S-Aggregate With this method we consider the potential differences in the actual profit when adopting the advice and when sticking to the user's initial choice (i.e., before receiving the advice). The idea is to provide the advice y that maximizes a weighted sum of these differences over all possible world sates (nature states). Formally, denoting the profit of choosing A when the world state turns to be w by $V(A|w)$, the advisor provides the advice y that satisfies: $\arg\max_y \sum_w Beta(Norm(V(y|w) - V(I|w)))p(w)$, where $p(w)$ is the a priori probability that the state of nature will turn to be w, I is the user's initial choice and $Norm(V(y|w) - V(I|w))$ is the difference $V(y|w) - V(I|w)$ normalized to the interval $0 - 1$, so that it could be used as an input for the Beta cumulative distribution function.

The reason for using the Beta distribution for this purpose relates to the essence of the difference. In some sense, the difference $V(y|w) - V(I|w)$ can be correlated with the notion of "regret". Most people regret when they realize an alternative decision would bring them better profit, retrospectively. The meaning of regret is that people care not only on what they actually receive, but also what they could have received had they chosen the alternative option, so their current utility is actually also a function of former possible decisions (Zeelenberg 1999). Prior work has dealt extensively with the way people take the feeling of regret into account when facing a decision making situation (Bleichrodt, Cillo, and Diecidue 2010). It was found to be a more complex emotion than the basic ones like anger, fear or happiness, as people usually tend to compare their decision's outcome to what they could have gotten had they chosen differently. Moreover, regret is a negative emotion, as large intensities of regret are weighted disproportionally heavier than small ones.While regret is, by definition, always positive (or zero), the difference $V(y|w) - V(I|w)$ in our case can also be negative. Meaning that we are not measuring the user's regret from her decision given a specific state of nature per-se. Instead, we are measuring the user's "regret" should she had switched her original choice and used the advice provided. In that sense, switching to adopting the advice provided can be accompanied by either a relative gain or relative loss.

The characterization of the differences as gains and losses calls for some adjustment in the way these values are aggregated. It is well known that people are highly affected by the nature of outcomes (gains or losses). In particular, people are greatly (negatively) affected by losses compared to (the positive effect of) gains. In Prospect theory, it is customary to use a (subjective) S-shaped value function (Tversky and Kahneman 1981) combined of a concave function for gains and convex function for losses, as shown in Figure 1.

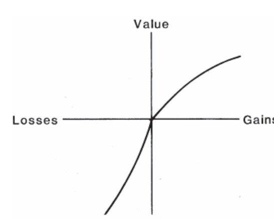

Figure 1: A hypothetical valuation function for gains and losses (Kahneman and Tversky 1979).

This latter behavior can be captured by the Beta cumulative distribution function —with a proper use of the α and

[4] A somehow similar idea of "step by step" advice-giving is used for teaching—starting with small hints and gradually increases the level of detail of the advice given (Watanabe et al. 2003).

β parameters (different sets for valuing positive and negative differences) we manage to assign substantial (negative) value to cases where the user's initial plan turns to be better than the advisor's. For the opposite case (advisor's advice turned to result in a better outcome than the user's initial plan) the value assigned still monotonically increases in the difference, yet in a slower increase rate compared to the case of losses. Using Beta functions with parameters $\alpha = 1$ and $\beta = 4$ for valuing (normalized) positive differences and $\alpha = 0.7$ and $\beta = 4$ for valuing (normalized) negative differences we obtain a valuation function that aligns with the one suggested by by Kahneman and Tversky (1979), as given in Figure 1. This way we actually put more emphasis on the chance that the advice will result in a better outcome compared to the user's initial plan, rather than on the magnitude of (or the average value of) the improvement achieved. This latter property aligns with evidence given in recent work for people's preference of decision strategy that favors winning most of the time over the one that yields higher final value (Hills and Hertwig 2010).

Experimental Framework

For our experiments we used a common repeated decision making setting captured by the *Car Purchasing Game* (CPG) (also known as "Acquiring a Company game") that was originally introduced by Samuelson and Bazerman (1985). In this game there are two players: the *seller* of a car and a *buyer* who is interested in buying it. The actual condition of the car, and consequently its value, denoted v ($v > 0$), is privately held by the seller. The buyer is only acquainted with the underlying a priori probability distribution function of the car's worth, denoted $f(v)$. If buying the car, the buyer will be able to improve its condition (e.g., she knows a gifted mechanic) by a factor x ($x > 1$), making its worth $x \cdot v$. The buyer needs to come up with a "take it or leave it" offer (denoted O) to the seller, meaning that if O is above v the seller will accept it, selling the car to the buyer at that price, and otherwise reject, leaving the buyer with a zero profit. The buyer's goal in this game is to maximize her expected profit, defined as the car's worth v times x minus the payment O.

The buyer's expected profit can be explicitly expressed as $\int_{v=0}^{O}(x \cdot v - O)f(v)dv$.[5] The expected-profit-maximizing offer (to which we refer as "optimal" onwards), denoted O^*, can be trivially extracted by taking the first derivative of the latter expression and equating it to zero. In particular, when the car's value is uniformly distributed over some interval $(0, V)$, as used in all prior work relying on CPG-like settings, then $O^* = 0$ for any $x \leq 2$ and $O^* = V$ otherwise.[6]

While the problem is common and most people face it (or a similar one) quite often, its solution for the uniform distribution function is highly non-intuitive to people—the optimal strategy depends solely on x, and shifts from one extreme to another at the point $x = 2$ using a step function-like pattern. It suggests that the car should not be purchased at all for cases where $x < 2$, even though the buyer values it more than the seller (and even for the case where $x = 2 - \epsilon$, i.e., the buyer values it almost twice its value to the seller). Similarly, when $x > 2$ one should offer the maximum possible worth of the car, even though it is possible that the car is worth way less. In particular, the transition between the two extremes (avoid purchasing and offer the maximum possible) at $x = 2$ is confusing for people. Indeed, prior research has given much evidence to a substantial deviation in the offers made by human subjects in this setting from the optimal ones. For example, studies that used $x = 1.5$ have shown that most offers people make lie between the expected value ($V/2$) and the ex-ante expected value of the item ($1.5 \cdot \frac{V}{2}$) (Carroll, Bazerman, and Maury 1988). A possible explanation for this is that participants simplify their decision task by ignoring the selective acceptance of the seller, as if the seller does not have more information than they have. Similarly, for a variety of x values within the range $2 - 3$ it has been shown that people make offers that are substantially lower than V. Furthermore, it has been shown that even when providing people with the optimal offers they tend not to adopt them (Elmalech et al. 2015). Unlike the general acceptance among experimental economists that optimal behavior should not necessarily be expected right away, but is likely to evolve through the process of learning and adjustment (Kagel 1995), participants in our settings were reported to exhibit a strong persistence in suboptimal behavior, even when experiencing with the task for the purpose of fostering learning (Grosskopf, Bereby-Meyer, and Bazerman 2007; Carroll, Bazerman, and Maury 1988).

The fact that people are known to act suboptimally and fail to adopt the optimal solution in CPG makes it a perfect framework for testing our advice generation methods. Furthermore, the fact that it has been extensively used in prior work enables us validating our experimental findings through the comparison of a "no-advisor" control group to prior reported results.

Experimental Design

We implemented the CPG as a java-script web-based application such that participants could interact with the system using a relatively simple graphic interface, thus facilitating interactions with a variety of people (see screenshots in Figure 2). The car's worth a priori distribution was set to be uniform between 0 and 1000 (i.e., $V = 1000$). From the user-interface point of view, the advice (i.e., the suggested offer) was presented to participants by a virtual advisor. Participants did not receive any information related to the nature of the advisor, and the advice they received used the text "The best advice I can give you is: offer to the seller K dollars" (where K is the advice in the current game). The interface was designed such that we can run experiments with and without an advisor. When using an advisor, after introducing the user the x value for the game, she is first asked to input her initial offer (see Figure 2(a)). Then the advisor provides the advice and the user can either adopt the advice received

[5] The calculation goes over all possible v values for which the offer is accepted. For all other values the profit is zero.

[6] Intuitively, in this case the expected worth of the car, given that an offer $O \leq V$ was accepted, is $O/2$ and therefore the expected profit is $(O/V)(x \cdot O/2 - O) = \frac{O^2}{V}\frac{x-2}{2}$, which is maximized for $O = 0$ when $x \leq 2$ and $O = V$ otherwise.

or use the offer she initially intended to use before receiving the advice (see Figure 2(b)). Once the final offer was set, the user became acquainted with the actual worth of the car v, and consequently whether or not it was accepted (see Figure 2(c)). The system also calculated and displayed to the user the actual benefit obtained based on the offer made and the potential benefit if she had chosen the other offer (either the advisor's suggestion, in case of sticking with the original user's offer, or the original offer if switching to the advisor's suggestion).

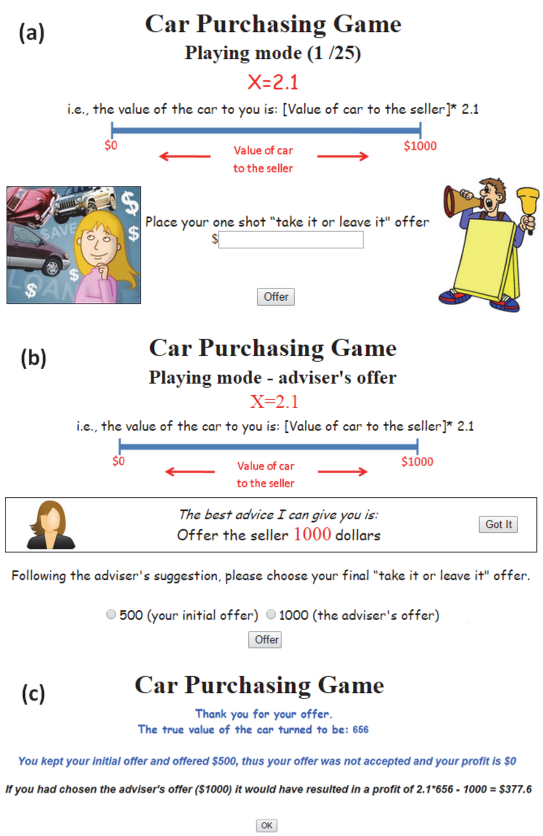

Figure 2: Screen-shots of the CPG interface.

We implemented both our suboptimal advisors, for the CPG. The S-Gradual advisor was implemented with an initial advice of \$300 for $x < 2$ and \$700 for $x > 2$.[7] Each subsequent advice further converged towards the expected-profit-maximizing advice by \$30. This enabled reaching the highly non-intuitive yet optimal advices of \$0 and \$1000 (for $x < 2$ and $x > 2$, respectively) within 10 rounds (for each x value type). The S-Aggregate method was implemented using the possible car's worth values as possible world states (w), where I is the user's initial offer as received in the system. The Beta function parameters used were ($\alpha = 1, \beta = 4$) and ($\alpha = 0.7, \beta = 4$) for positive and negative differences, respectively, for the reasons given above. In addition, we have implemented two advisors, to be used as a benchmark for evaluating the performance of our two suboptimal advice provisioning methods. The first is the advisor suggested by Elmalech et al. (2015), denoted "S-Elmalech" onwards. This advisor generates suboptimal advices for CPG-like settings however was initially designed and tested for non-repeated interaction. The second is an advisor that always suggests the optimal offer. i.e., \$0 in case $x < 2$ and \$1000 otherwise, denoted "Optimal" onwards.

Participants were recruited and interacted through Amazon Mechanical Turk (AMT) which has proven to be a well established method for data collection of tasks which require human intelligence to complete (Paolacci, Chandler, and Ipeirotis 2010). To prevent any carryover effect a "between subjects" design was used, assigning each participant to one experiment only, which was randomly selected out of five possible treatments: (a) playing the game without an advisor; (b) playing the game with the Optimal advisor; (c) playing the game with the S-Elmalech advisor; (d) playing the game with our S-Gradual advisor; and (e) playing the game with our S-Aggregate advisor.

The compensation for taking part in the experiment was composed of a show-up fee (the basic "HIT") and also included a bonus, which was linear in the participant's average profit over all games played, in order to encourage thoughtful participation. Each participant received thorough instructions of the game rules, the compensation terms and her goal in the game. Then, participants were asked to engage in practice games until stating that they understood the game rules (with a strict requirement for playing at least three practice games). Prior to moving on to the actual games, participants had to correctly answer a short quiz, making sure they fully understand the game and the compensation method. Finally, participants had to play a sequence of 25 CPGs, each differing in the value of x used.[8] In order to have a better control over the experiment, we used 10 randomly generated sequences of x values (25 values in each) and each experiment was randomly assigned one of these sequences. Similarly, the actual car's worth v associated with each of the ten x sequences in each experiment was taken from one of ten pre-drawn sets of 25 values within the range $0 - 1000$.[9] For each game the system stored the offer originally made, the advice given, the offer eventually picked, the car's true worth and the profit with both offers (the user's original offer and the advice received). While participants did not receive any information related to the advisor and the nature of its advice, they were told that they will be using the same advisor throughout all the 25 games.

In order to evaluate the user satisfaction with the different advisors, we asked users to answer the following three questions at the end of the experiment:

1. *"Overall, how satisfied were you with the advisor?"*
2. *"Based on your experience with the advisor, would you recommend to a friend using it?"*

[7]The choice of these initial values is based on Elmalech et al. (2015), where these were found to be the most commonly used by people when not given any advice, in a non-repeated setting.

[8]The choice of 25 games was made in order to push people to use expected-benefit maximization strategy, as discussed above.

[9]See supplementary material in the authors' web-site for the full set of values used.

3. *"If you had to choose between playing with or without the advisor given, what would you choose?"*

Each of the three questions relate to user satisfaction from a slightly different aspect. For example, the answer to the second question reflects user's loyalty and there is a direct link between loyalty and user satisfaction (Hallowell 1996).

Overall, we had 500 participants taking part in our experiments, 100 for each treatment, each playing 25 games according to the above design. Participants ranged in age (21-70), gender (61% men and 39% women), education (25% with secondary education, 50% bachelor's degree, 17% masters) and nationality (63% from US, 30% from India), with a fairly balanced division between treatments.

Results

Results are presented in a comparative manner, according to the average profit, advice adoption rate and user satisfaction.

User Profit Figure 3 depicts the difference in the average profit (cross-participant) between the NoAdvice and the four other treatments (methods) used. It also includes the statistical significance of the difference between the different treatments (reporting only cases where the $p-value$ obtained is less than 0.1). The statistical test used is the Mann Whitney U-test (also known as the Wilcoxon rank-sum test) which is a nonparametric test of the null hypothesis that two samples come from the same population against the alternative hypothesis that one population tends to have larger values than the other (Nachar 2008). The advantage of this test in the context of our experimental design is that it can be applied on unknown distributions, contrary to $t-test$ which has to be applied only on normal distributions, and it is nearly as efficient as the t-test on normal distributions.

Based on Figure 3 we conclude that both our S-Gradual and S-Aggregate methods perform at least as good as all the other tested methods as far as user's profit is concerned. In particular, S-Gradual performs substantially better than all other methods that were checked (including S-Aggregate), suggesting an improvement of $55\%-71\%$ in average profit. Interestingly, the S-Elmalech method that was shown experimentally to result in the best profit in very similar yet non-repeated settings (Elmalech et al. 2015) is actually a poor choice once switching to a repeated setting, performing worse than all other methods (and substantially worse (statistically significant) than our S-Gradual method).

Advice Adoption Rate In our experiments we had 2500 advices given in each treatment (100 participants, each playing 25 games), hence the adoption-rate relates to the percentage of these advices that were adopted. The average adoption-rate in our experiments was $25\%, 35\%, 37\%$ and 38% for the Optimal, S-Elmalech, S-Aggregate and S-Gradual advisors, respectively. These results suggest that the advices of all three suboptimal advisors were generally more appealing to participants. The breakdown of the above averages according to the number of times the advice was adopted with each advisor, as well as additional insights, appear in the supplementary material in the authors' web-site.

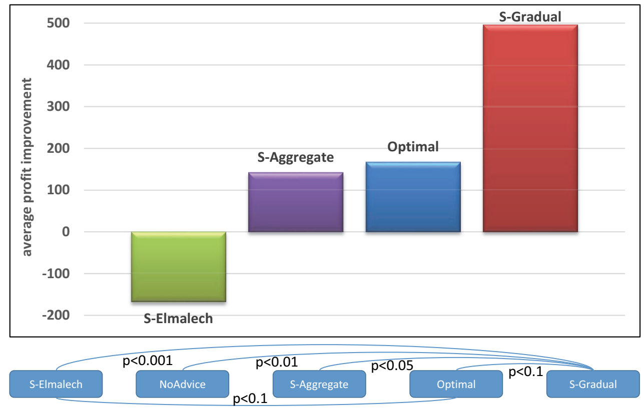

Figure 3: The difference in the average actual profit between each treatment and the NoAdvice treatment and statistical significance results (p-value).

Putting the acceptance rate with the different methods over the timeline, we observed the following patterns: (a) the acceptance rate with the Optimal advisor converges to 25%, hence it is likely to remain within this level in further interactions (i.e., beyond the 25 tested interactions for each participant); (b) the acceptance rate with the S-Elmalech advisor is stable over time, hence is also likely to remain within this level in additional interactions; (c) with the S-Aggregate advisor the acceptance rate was stable within the first 16 interactions, and then it monotonically increased in a relatively constant pace; and (d) with the S-Gradual advisor the acceptance rate is the highest (compared to all other methods) within the first few interactions (as the offers made at this stage are highly intuitive), then it decreases (as offers became less intuitive) and finally stabilizes within the last few interactions. Interestingly, the adoption rate level within the last few interactions with S-gradual is greater than the level obtained with the Optimal advisor despite providing the same advice (as by then S-Gradual advisor has finished its gradual transition towards the optimal advice). This can possibly indicate that the users find it more competent due to the fact that it provided more "reasonable" advices at first.

Offers Made In an attempt to better understand the source of improvement with the different methods compared to the NoAdvice case, we calculated the average initial offer users made within the first and last six games of the experiment. The results (brought in detail in the supplementary material in the authors' web-site) indicate that there is no statistically significant difference between the averages in early and later games for all advisors. Meaning that none of the advisors affected the user's understanding of the solution to the problem. The performance improvement achieved with our methods can therefore be fully attributed to the user's perception of the advisor's competence (which is also reflected in the increase in the adoption rates) and the value encapsulated in the advices given.

Furthermore, the distribution of offers for the NoAdvice treatment was used for validation purposes, as this is a setting similar to the ones originally used in prior work. The results are indeed quite similar. For example, the percentage of participants offering $600 and more when using $x = 1.6$ in our experiment is 36% compared to 38% in Samuelson et al. (1985) and 36% in Elmalech et al (2015).

User's satisfaction Figure 4 depicts the results of user satisfaction based on the three questions that users had to answer at the end of the experiment. The proportions given in graphs (a), (b) and (c) correspond to the percentage of people who felt generally satisfied with the advisor, would recommend it to a friend and stated that will use the advisor again, respectively. The figure also includes the statistical significance of the difference in the proportions between the different treatments (reporting only cases where the $p - value$ obtained is less than 0.1). The statistical test used is the Z-Test for two population proportions.

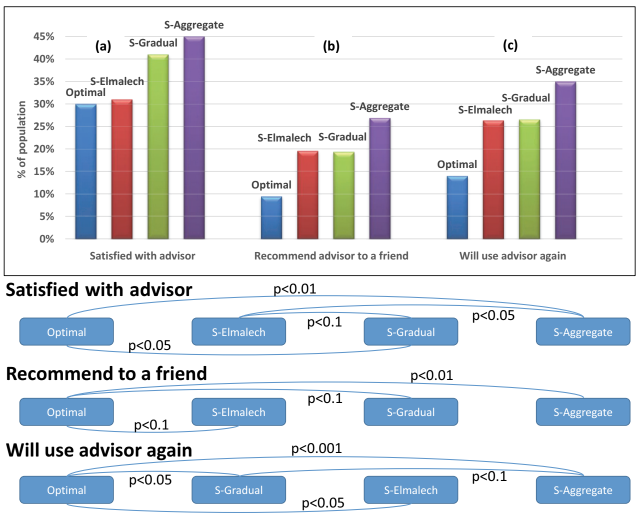

Figure 4: User satisfaction survey results.

Based on Figure 4 we conclude that both S-Gradual and S-Aggregate perform at least as good as all the other tested methods as far as user satisfaction is concerned. In particular, both methods perform substantially better (statistically significant) than all other methods that were checked in the general satisfaction question, with an improvement of $32\% - 50\%$ in that measure. In the two other questions both methods, as well as S-Elmalech, were found to be significantly better than the Optimal advisor, providing evidence to the strength of the suboptimal advising approach in general. Interestingly, the S-Elmalech method that was found to perform worse than the Optimal advisor as far as average profit is concerned, was found to result in a better user satisfaction (statistically significant for two of the questions). This strengthens our hypothesis that the average profit is not necessarily a sufficient measure for user satisfaction.

Discussion, Conclusions and Future Work

The (relative) failure of the Optimal advisor both in the average profit and user satisfaction fronts, as well the low acceptance rate it achieved compared to the other methods, provide a strong motivation for the development of advice provisioning methods that use suboptimal advice. The failure of the S-Elmalech method in the repeated version of the setting it was initially designed for is yet another evidence for the need to apply different designs whenever the advisor is used beyond a single interaction. The results reported in the former section suggest that both methods studied in this paper are effective in repeated advice provisioning settings and meet their design goals—the use of the S-gradual results in a significantly improved performance (a greater average profit in our experimental infrastructure) compared to all other tested methods, while keeping user satisfaction at least at the same level (or without a statistically significant reduction) as with the other methods. The use of the S-Aggregate results in a significantly improved user satisfaction compared to all other tested methods, while keeping average profit at least at the same level (or without a statistically significant reduction) as with the other methods. As our experimental results show, maximizing the actual profit and maximizing the user's satisfaction from the advisor are two different things with little correlation between them. The choice of which should be maximized and how to tradeoff the two is the agent designer's, hence the usefulness of having the two methods rather than one.

One encouraging finding related to S-aggregate is that its acceptance rate was found to monotonically increase over the last few interactions of the experiment. A high acceptance rate for itself is an important desirable property of an advisor, as argued throughout the paper. Another implication of this finding, however, is that the increase in the adoption rate of S-Aggregate is likely to translate to better profits over time, hence this method is likely to perform even better (than the current reporting) if extending the experiments beyond the 25 interactions. This latter phenomena does not recur in the other methods. A possible limitation of the S-aggregate method is that it requires knowing what the user is about to do by herself prior to receiving the advice. While this is inapplicable for some settings, there are many others where such information can be obtained, e.g., through user modeling, asking the user directly or simply delaying the advice until after observing the user's plan for the next move.

We note that, much like many other works in the area of advice provisioning and human-agent interaction in general, our experimental design uses only a single testbed (CPG). Obviously, testing the proposed methods in additional domain will strengthen the generality of the results. Having said that, we emphasize that the CPG is a well established infrastructure, that has been extensively used in prior research, and formerly reported results enabled us in this case to validate the experimental design used. Furthermore, the methods presented in this paper are general (i.e., do not relate to specific characteristics of CPG), and can be applied in various other situations. Also, the decision setting captured by the CPG is similar to the ones found in various other real life decision situations, e.g., investments in stocks, bonds and options.

An important direction for future work is the integration of methods for user classification in order to fit the advice provided to the specific user. Other possible extensions of this work include the integration of formal methods for estimating user's trust and confidence in the advisor's design.

Acknowledgments

This research was partially supported by the ISRAEL SCIENCE FOUNDATION (grant No. 1083/13) and the ISF-NSFC joint research program (grant No. 2240/15).

References

Azaria, A.; Kraus, S.; Goldman, C. V.; and Tsimhoni, O. 2014. Advice provision for energy saving in automobile climate control systems. In *AAAI*, 2954–2959.

Azaria, A.; Gal, Y.; Kraus, S.; and Goldman, C. 2015. Strategic advice provision in repeated human-agent interactions. *Autonomous Agents and Multi-Agent Systems* 1–26.

Azaria, A.; Kraus, S.; and Richardson, A. 2013. A system for advice provision in multiple prospect selection problems. In *RecSys*, 311–314.

Bharati, P., and Chaudhury, A. 2004. An empirical investigation of decision-making satisfaction in web-based decision support systems. *Decision support systems* 37(2):187–197.

Bleichrodt, H.; Cillo, A.; and Diecidue, E. 2010. A quantitative measurement of regret theory. *Management Science* 56(1):161–175.

Carroll, J. S.; Bazerman, M. H.; and Maury, R. 1988. Negotiator cognitions: A descriptive approach to negotiators' understanding of their opponents. *Organizational Behavior and Human Decision Processes* 41(3):352–370.

Elmalech, A.; Sarne, D.; Rosenfeld, A.; and Shalom Erez, E. 2015. When suboptimal rules. In *Proceedings of AAAI'15*, 1313–1319.

Elmalech, A.; Sarne, D.; and Grosz, B. J. 2015. Problem restructuring for better decision making in recurring decision situations. *AAMAS* 29(1):1–39.

Grosskopf, B.; Bereby-Meyer, Y.; and Bazerman, M. 2007. On the robustness of the winner's curse phenomenon. *Theory and Decision* 63(4):389–418.

Hajaj, C.; Hazon, N.; and Sarne, D. 2014. Ordering effects and belief adjustment in the use of comparison shopping agents. In *Proc. of AAAI*, 930–936.

Hajaj, C.; Hazon, N.; and Sarne, D. 2015. Improving comparison shopping agents competence through selective price disclosure. *Electronic Commerce Research and Applications*.

Hallowell, R. 1996. The relationships of customer satisfaction, customer loyalty, and profitability: an empirical study. *International Journal of Service Industry Management* 7(4):27–42.

Hills, T., and Hertwig, R. 2010. Information search in decisions from experience: do our patterns of sampling foreshadow our decisions? *Psych. Sci.* 21(12):1787–1792.

Jessup, R. K.; Bishara, A. J.; and Busemeyer, J. R. 2008. Feedback produces divergence from prospect theory in descriptive choice. *Psychological Science* 19(10):1015 – 1022.

Kagel, J. 1995. *Auctions: a survey of experimental research*. Princeton: Princeton University Press. 501–513.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–291.

Klos, A.; Weber, E. U.; and Weber, M. 2005. Investment decisions and time horizon: Risk perception and risk behavior in repeated gambles. *Management Sci.* 51(12):1777–1790.

Lemon, K.; Barnett White, T.; and Winer, R. 2002. Dynamic customer relationship management: Incorporating future considerations into the service retention decision. *Journal of Marketing* 66(1):1–14.

Lesser, L. M. 1999. Exploring the birthday problem with spreadsheets. *The Mathematics Teacher* 407–411.

Martin, A. D.; Quinn, K. M.; Ruger, T. W.; and Kim, P. T. 2004. Competing approaches to predicting supreme court decision making. *Perspectives on Politics* 2(4):761–767.

Nachar, N. 2008. The mann-whitney u: A test for assessing whether two independent samples come from the same distribution. *Tutorials in Quantitative Methods for Psychology* 4(1):13–20.

Paolacci, G.; Chandler, J.; and Ipeirotis, P. 2010. Running experiments on amazon mechanical turk. *Judgment and Decision Making* 5(5):411–419.

Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior & Organization* 3(4):323–343.

Ricci, F.; Rokach, L.; Shapira, B.; and Kantor, P. B., eds. 2011. *Recommender Systems Handbook*. Springer.

Rochlin, I., and Sarne, D. 2013. Information sharing under costly communication in joint exploration. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*, 847–853.

Rochlin, I., and Sarne, D. 2014. Constraining information sharing to improve cooperative information gathering. In *Proceedings Of AAMAS*, 237–244.

Rochlin, I.; Sarne, D.; and Mash, M. 2014. Joint search with self-interested agents and the failure of cooperation enhancers. *Artificial Intelligence* 214:45–65.

Rosenfeld, A., and Kraus, S. 2012. Modeling agents based on aspiration adaptation theory. *JAAMAS* 24(2):221–254.

Rosenfeld, A., and Kraus, S. 2015. Providing arguments in discussions based on the prediction of human argumentative behavior. In *Proceedings of AAAI'15*, 1320–1327.

Rosenfeld, A.; Zuckerman, I.; Azaria, A.; and Kraus, S. 2012. Combining psychological models with machine learning to predict people's decisions. *Synthese* 189(1):81–93.

Rosenfeld, A.; Agmon, N.; Maksimov, O.; Azaria, A.; and Kraus, S. 2015. Intelligent agent supporting human-multi-robot team collaboration. In *IJCAI*, 1902–1908.

Samuelson, W. F., and Bazerman, M. H. 1985. Negotiation under the winner's curse. *Research in experimental economics* 3:105–38.

Starmer, C. 2000. Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk. *Journal of Economic Literature* 38(2):332–382.

Tversky, A., and Kahneman, D. 1981. The framing of decisions and the psychology of choice. *Science, New Series* 211:453–458.

Watanabe, H.; Takai, K.; Arai, M.; and Takei, S. 2003. Case-based adviser for near-miss programs. *Artificial Intelligence in Education* 149–156.

Zeelenberg, M. 1999. Anticipated regret, expected feedback and behavioral decision making. *Journal of Behavioral Decision Making* 12(2):93–106.

Zhang, Y.; Bellamy, R. K.; and Kellogg, W. A. 2015. Designing information for remediating cognitive biases in decision-making. In *Proceedings of CHI '15*, 2211–2220.