# The influence of task contexts on the decision-making of humans and computers.

# The Influence of Task Contexts on the Decision-making of Humans and Computers

Ya'akov Gal[1,2], Barbara Grosz[2], Avi Pfeffer[2], Stuart Shieber[2], and Alex Allain[3]

[1] MIT Computer Science and Artificial Intelligence Laboratory, Cambridge MA 02139
[2] Harvard School of Engineering and Applied Sciences, Cambridge MA 02138
[3] Liquid Machines, Inc., Waltham MA 02451

**Abstract.** Many environments in which people and computer agents interact involve deploying resources to accomplish tasks and satisfy goals. This paper investigates the way that the context in which decisions are made affects the behavior of people and the performance of computer agents that interact with people in such environments. It presents experiments that measured negotiation behavior in two different types of settings. One setting was a task context that made explicit the relationships among goals, (sub)tasks and resources. The other setting was a completely abstract context in which only the payoffs for the decision choices were listed. Results show that people are more helpful, less selfish, and less competitive when making decisions in task contexts than when making them in completely abstract contexts. Further, their overall performance was better in task contexts. A predictive computational model that was trained on data obtained in the task context outperformed a model that was trained under the abstract context. These results indicate that taking context into account is essential for the design of computer agents that will interact well with people.

## 1 Introduction

Technology has opened up vast opportunities for computer agents to interact with people in such increasingly diverse applications as online auctions, elderly care systems, disaster relief operations, and system administrator groups [1, 10]. While these applications differ broadly in size, scope, and complexity, they are similar in that they involve people and computers working together in *task settings*, in which the participants fulfill *goals* by carrying out *tasks* requiring the use of *resources*. Participants may need to cooperate, negotiate, or perform other group actions in order to achieve the goals, requiring their reasoning about the potential and likely behaviors of other participants. For computer agents to interact successfully with people in such mixed human-computer task settings, they need to meet people's expectations of teammates.

For example, in the domain of care of elderly patients, the physical challenges and health problems of this population typically require a team of caretakers—not only doctors and nurses, but also home health aides, housekeepers, family members. Current medical care depends on computer systems for scheduling and tracking prescriptions; computers, of a very small scale, are also key elements of pacemakers and other implantable medical devices. Thus, the agents involved in elder care are both human and computer-based; they come from different organizations and have different roles. As the computer agents involved in such care become more sophisticated and more of them become connected to the care of a single individual, the need for abilities to coordinate and work as team members will become important.

In designing computer agents for such settings, it is thus important to understand the decision-making strategies people deploy when they interact with others and to evaluate various computational strategies for interacting with people. Formally modeling the behavior of people, and in particular their decision-making behaviors, raises significant challenges for computer-agent design.

To investigate the influence of task contexts on decision-making, we deployed a conceptually simple but expressive game called Colored Trails (CT) [4]. CT explicitly manifests goals, tasks, and resources in a way that is compelling to people, yet abstracts away from a complicated underlying domain. By embedding decision-making within a task context, CT enables investigators to focus on people's decision-making strategies, rather than specifying and reasoning about individual domain complexities.

CT differs significantly from the highly abstracted settings typically used in behavioral economics, such as decision trees or normal form tables. These forms completely hide the underlying relationship between tasks, goals, and resources and fully specify payoffs for players from potential strategies. We call this abstract representation a *table context*. Game-theoretic tools can be applied in such games to provide an idealized notion of appropriate decision-making behavior. The decisions engendered by CT games can also be described as a table of payoffs, enabling to contrast between task and table contexts use to embed the same decision.

We analyzed people's behavior in terms of various social criteria, for which we give a precise definition in terms of the CT game. We show that people presented with identical decision-making problems in the task context and the table context perform strikingly differently, both qualitatively and quantitatively. When making decisions in the task context, people are more helpful, less competitive and less game-theoretic than when making decisions in the table context. Surprisingly, the results also indicate that the task context improves people's overall performance.

To evaluate the effects of these differences for computer agents that interact with people, we trained predictive models on data obtained in both types of contexts. The models explicitly represented social factors that have been shown to affect people's behavior [3]. Most importantly, the model trained on data obtained in the task context outperformed the model trained on data obtained in the table context. In addition, overall performance was better when the context was task-oriented, rather than payoff-oriented.

For designers of intelligent agents, the important lesson of these experiments is that the design of computer agents that will operate in mixed human-computer settings must consider how the decisions presented to people will be contextualized and reflect the human decision-making process in that context, not merely in a purely idealized (even if theoretically equivalent) manner. As much as we might like it, there is no way for computer agents to escape into pure game theory when participating in mixed systems.

## 2 Empirical Methodology

This section describes the two types of context, task context and table context, we investigated and the experiments conducted in those settings.

In the *task* context, a 2-player CT game was played on a 4x4 board of colored squares with a set of chips. One square on the board was designated as the goal square. Each player's icon was initially located in a random, non-goal position. To move to an adjacent square a played needed to surrender a chip in the color of that square. Players were issued four colored chips. They had full view of the board and each others' chips, and thus they had complete knowledge of the game situation.

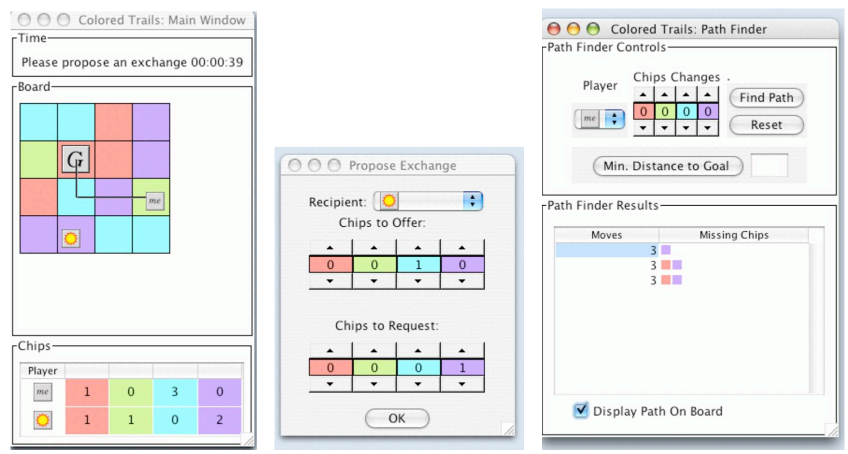
Players were designated one of two roles: *proposer* players could offer some subset of their chips to be exchanged with some subset of the chips of responder players; *responder* players could in turn accept or reject proposers' offers. If no offer was made, or if the offer was declined, then both players were left with their initial allocation of chips. Chip exchanges were enforced by the game controller: after the negotiation ended, both players were automatically moved as close as possible to the goal square.

The scoring function for players depended solely on their own performance: 100 points for reaching the goal; 10 points for each tile left in a player's possession; 15 points deducted for any square in the shortest path between player's final position and the goal-square. These parameters were chosen so that getting to the goal was by far the most important component, but if an player could not get to the goal it was preferable to get as close to the goal as possible. The score that each player received if no offer was made was identical to the score each player received if the offer was rejected by the deliberator. We refer to this score as the *no negotiation alternative* and to the score that each player received if the offer was accepted by the deliberator as the *proposed outcome* score.

Snapshots of the CT GUI of one of the games used in the experiment is shown in Figure 1. The Main Window panel, shown in Figure 1a, includes the board game, the goal square, represented by an icon displaying the letter *G*, and two icons, "me" and "sun", representing the location of the two players on the board at the onset of the game.[4] The bottom part of the Main Window panel, titled "chips", shows the chip distributions for the players. In the game shown here, both players lack sufficient chips to get to the goal square. A proposer uses the Propose Exchange panel, shown in Figure 1b, to make an offer to a responder. The Path Finder panel, shown in Figure 1c, provides decision support tools to be used during the game. It displays a list of path suggestions to the goal, the missing chips required to fulfill each path, and the best position the agent can reach relative to its scoring function. Agents can view this information for the chip set that is currently in their possession, or for any hypothetical chip set for each of the players.

The *table* context consisted of a completely abstract representation of a CT game as a list of potential offers that could be selected by the proposer player. Each offer was represented as a pair of payoffs for the proposer and the responder. Figure 2 shows a snapshot of a game in this representation as seen from the point of view of a proposer player. Each cell in the table represents an offer, and selecting a cell corresponds

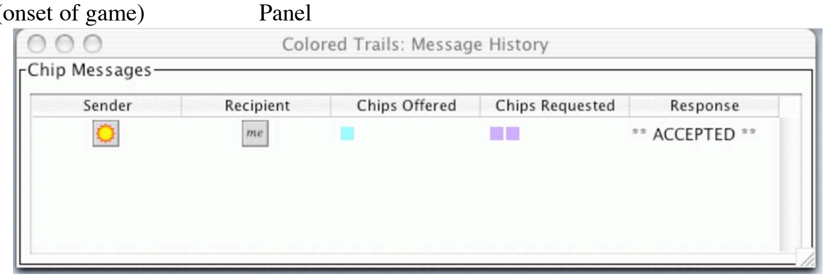
[4] CT colors have been converted to grey scale in this figure.




(a) Main Window Panel (onset of game)

(b) Propose Exchange Panel

(c) Path Finder Panel




(d) Message History Panel

Fig. 1: Snapshots of Interaction in a Task Context

to choosing the offer associated with its payoffs. One of the cells represents the no-negotiation alternative, which is presented as the default outcome of the interaction.

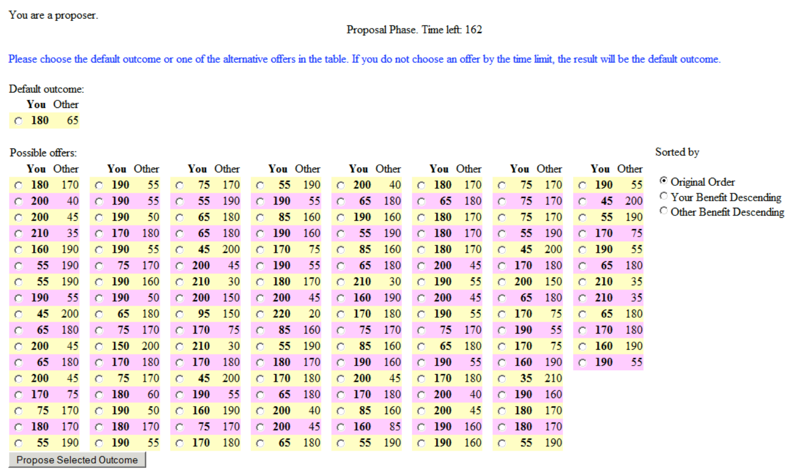



Fig. 2: Snapshot of an Interaction in a Table Context

A total of 32 subjects participated in the experiment, equally divided between the two conditions. They interacted with each other for 96 rounds. Participants in the task condition interacted with each other using the CT environment, whereas those in the table condition interacted with each other using the payoff matrix representation. Participants only interacted with others in their condition group; they were not provided any information about each other. In both conditions, participants were compensated in a manner that depended solely on their individual scores, aggregated over all rounds of interaction.

For each CT round that was played in the task condition, an equivalent round was played in the table condition, in the sense that the payoff pair at the intersection of each row and column represented the score in the CT round for the corresponding offer and response. For example, the payoff matrix shown in Figure 2 is equivalent to the CT game shown in Figure 1.

## 3 Results and Analysis

We use the term *table proposers* and *task proposers* to refer to the participants that were designated with the proposer role in the table or task condition respectively and similarly for the responder role. We use the term *offer benefit* to refer to the difference between the proposed outcome for an offer and the no-negotiation alternative score of the round. We measured proposers' behavior in terms of two features: The degree to which proposers were *selfish* or *helpful* was defined in terms of the average offer benefit they proposed for themselves or for responders, respectively; the degree to which proposers were *competitive* was defined in terms of the difference between the average offer benefit they proposed for themselves and the offer benefit they provided to responders. Although we have given a psychological interpretation to these features, we do not imply that they are independent. For example, proposers can exhibit both a degree of selfishness and a degree of helpfulness based on the average benefit of their offers.

### 3.1 The Effect of Contexts on Human Behavior

Table 1 presents the average offer benefit to participants in both task and table condition for each role designation. Table proposers offered significantly more benefit to

Table 1: Average Benefit of Offer

| | Offer Benefit to | | Num. |
|---|---|---|---|
| | Proposer | Responder | acceptances |
| Task | 82.3 | **47.6** | 62 (77%) |
| Table | **98** | 36 | 69 (77%) |

themselves than did task proposers (t-test $p < 0.05$). Also, table proposers offered significantly less benefit to table responders than task proposers offered to task responders (t-test $p < 0.01$). Thus, the task context had the effect of making proposers more helpful and less selfish when interacting with responders.

The difference between the average offer benefit to proposers and to responders is positive in both conditions (t-test $p < 0.05$). Although in both conditions proposers are competitive, the offer difference was larger in the table condition than in the task condition (t-test $p < 0.05$). Thus, on average table proposers were more competitive than task proposers. We hypothesized that table proposers made competitive offers more often than did task proposers. To test this hypothesis, we performed a within-round comparison of the offer benefit in both conditions. Table 2 presents the number of rounds in which the difference between the proposed benefit for proposers and responders was positive (column "Proposer > Responder") and the number of rounds in which this difference was negative (column "Proposer < Responder" ). As shown by the table, table proposers made offers that benefited themselves over responders significantly more often than task proposers (chi-square $p < 0.05$). These results confirm that table proposers are more likely to be competitive than proposers.

Table 2 also shows that 62% of all offers made by table proposers benefited *themselves* more than table responders, while 60% of all offers made by task proposers benefited task *responders* more than themselves (chi-square $p < 0.05$). This striking result indicates that task proposers were helpful more often than they were selfish, whereas table proposers were selfish more often than they were helpful.

Having established that the context in which decisions are made affected the behavior of proposers, we investigated whether it affected the behavior of responders. It is

Table 2: Frequency of Competitive Offers

| | Proposer > Responder | Proposer < Responder |
|---|---|---|
| Task | 26 (27%) | **51** (60%) |
| Table | **60** (62%) | 24 (28%) |

more difficult to perform within-round comparisons of responder behavior across task and table conditions, because the decision of whether to accept or reject an offer depends on the exchange offered by proposers. For the same round, this exchange may be different for task and table conditions. As shown in Table 1, there was no difference in the ratio of exchanges accepted by responders (77%) between conditions. However, this result does not mean that responders were not affected by context; as also shown in Table 1, they were responding to exchanges that were more helpful to them in the task condition. We expected this pattern to hold for accepted offers as well; thus, we expected that the offers that were *accepted* by responders were more helpful to them in the task condition than in the table condition.

Table 3: Average Benefit for Accepted Exchanges

| | Proposer | Responder | Total |
|---|---|---|---|
| Task | 79.5 | **56.4** | 135.9 |
| Table | **85.6** | 40.7 | 126.3 |

Table 3 shows the exchange benefit to proposers and responders averaged over all accepted proposals, as well as the total accumulated benefit in each condition. The benefit to responders from accepted proposals was significantly higher in the task condition than in the table condition, and conversely for the proposers (t-test $p < 0.05$). These results indicated that task responders outperformed table responders, whereas table proposers outperformed task proposers. Interestingly, as the rightmost column shows, the total performance (combined proposers and responders scores) was higher in the task condition than in the table condition. The benefit for accepted exchanges is a measurement of performance, because the outcome of each round of interaction was fully determined by the action of the responder (t-test $p < 0.1$). Although this result was not significant at the $p < 0.05$ confidence interval, the trend it indicates suggests that task context has a positive effect on the combined performance of participants.

To compare between the benefits of proposed and accepted exchanges, we plotted the average benefit to proposer and responder from these offers in both conditions, as shown in Figure 3. We define the *discrepancy* between two offers to be the Euclidean distance between the two points representing the benefits of the offers to proposers and responders. As apparent from the figure, the discrepancy between proposed and accepted offers was significantly smaller in the task condition than in the idealized condition (t-test $p < 0.05$). This result suggests that on average, task proposers were more accurate at estimating the offers that were likely to be accepted by responders.

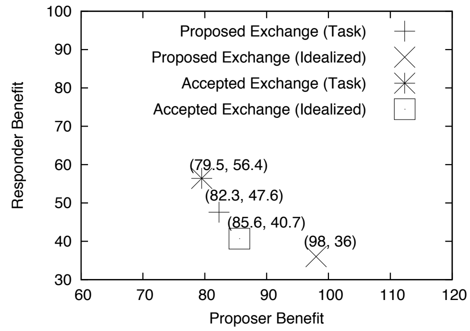



Fig. 3: Benefit for Proposed Exchanges vs. Accepted Exchanges

Also shown in Figure 3 is that in both conditions, accepted offers were more beneficial to responders than proposed offers; also in both conditions, accepted offers were less beneficial to proposers than proposed offers. This result suggests that responders expected proposers to be more helpful and less selfish in both conditions; this aligns with our findings that players were competitive across task and idealized contexts. However, the difference between the benefit to responders from proposed and accepted offers was significantly greater in the task condition than in the idealized condition. Similarly, the difference between the benefit to proposers from proposed and accepted offers was significantly greater in the idealized condition than in the task condition. This implies that in the idealized condition, responders expect proposers to be less selfish, while in the task condition, responders expect proposers to be more helpful. A possible explanation is that the task context induced responders to expect more help from proposers than the idealized context.

### 3.2 Discussion of Alternative Explanations

To address the question of whether the difference in behavior can be explained by the lack of an explicit representation of payoff in the task condition, we ran an experiment that used the CT game, but allowed subjects to view the payoffs for potential offers for all players. This intermediate representation preserves the task context as well as displaying the payoff function for both players. Results using the same set of games as in the original experiment show that there was no significant difference in the average benefit allocated to proposers and responders in this intermediate representation than in the task condition.

In addition, we ruled out the effect of cognitive demands on subjects by including decision support tools for both modes of decision representation. In the CT game, subjects could use the PathFinder panel, shown in Figure 1c to query the system for suggestions about the best paths to take given any hypothetical chip distribution. When presented with a table of payoffs in the table condition, subjects could sort the table by their own, or the others' benefit. In this way, subjects were allowed to focus on the interaction rather than on the cognitive complexity of the decision-making.

### 3.3 Comparison with Game Theoretic Strategies

We now turn to a comparison between the offers that were made in each condition and the offers dictated by the exchange corresponding to the Nash equilibrium strategy. We use the term *NE exchange* of a round to refer to the exchange prescribed by the Nash equilibrium strategy profile for the round. This exchange offers the maximum benefit for the proposer, out of the set of all of the exchanges that offer non-negative benefits to the responder. In our scenarios, the NE exchange generally amounted to selfish, unhelpful, competitive offers.

We expected table proposers to be more likely to offer NE exchanges than task proposers. Table 4 shows the number of NE offers made by proposers in both conditions. The proportion of NE offers was significantly higher in the table condition (59%) than in the task condition (15%) (chi-square $t < 0.01$).

Table 4: Frequency of Nash Equilibrium Offers

| | Num.. offers |
|---|---|
| Task | 13 (15%) |
| Table | **57** (59%) |

To compare the extent to which the exchanges made by proposers in the two type of contexts differed from the NE exchange, we plotted the average benefit offered by NE exchanges and by proposed exchanges for both task and table conditions, as shown in Figure 4.

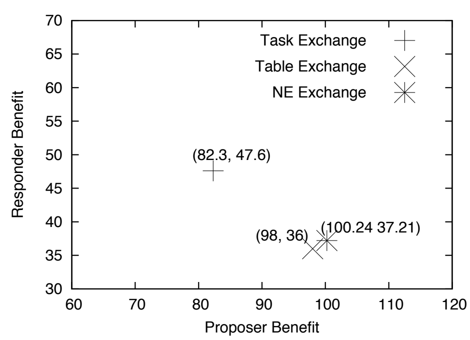



Fig. 4: Benefit from Proposed Exchanges vs. NE Exchanges

The difference between the average benefit to responders from the NE offer and the average proposed exchange was close to zero in the table condition, and large and positive in the task condition (t-test $p < 0.05$). Similarly, the difference between the benefit to proposers from the NE offer and the average proposed exchange was close to zero in the table condition, and large and negative in the task condition (t-test $p < 0.05$). The Euclidean distance between the two points representing the NE benefit to proposers and responders was significantly larger in the task condition than in the table condition. In fact, there was no statistically significant difference between offers in the table condition and NE offers. These results are strikingly significant, showing that participants who make decisions in the table condition are more likely to follow the game-theoretic paradigm.

There is a discrepancy between these findings and those of the behavioral economic studies, which show that people do *not* generally adhere to game theoretic equilibria, and display variance within their play. Several differences between the structure of the negotiation scenario used in our experiments and the games traditionally used in behavioral economics may explain this difference. First, our scenario presented participants with some guaranteed reward (the no-negotiation alternative) if agreement was not reached at the end of the interaction. Traditional behavioral economic games do not provide such reward. (For example, if no agreement is reached in the ultimatum game, both players end up empty handed.) It is possible that in table contexts, proposers saw fit to make selfish exchanges, because they could always fall back on their guaranteed outcome if that offer was rejected.[5] Second, each interaction in our experiment varied which player needed the other to get to the goal. In some rounds, both players were mutually dependent on each other. In traditional behavioral economic experiments, players' dependencies are static. A possible hypothesis, worthy of further investigation, is that table participants were more likely to follow game theoretic equilibria in one type of dependency but not in others.

## 4 The Effect of Contexts on Learner Agents

This section presents results that indicate the effects of task contexts on the performance of computer systems that learn a model of people's negotiation behavior. Using the data collected from the task and table contexts, we trained a computational model for predicting the actions of human proposers. We adopted the model proposed by Gal and Pfeffer [2] for predicting people's bidding behavior in multi-attribute negotiation. In this model, proposals are generated by converting people's utility functions into a stochastic function that assigns a probability to each potential exchange at each round of interaction.

At each round of interaction $k$, the inputs to the model were $NN_P^k$ and $NN_R^k$, the no-negotiation alternative scores for the proposer and responder, and $PO_P^k(x)$ and $PO_R^k(x)$, the proposed outcome scores for the proposer and responder for a potential exchange $x$. We omit the superscript $k$ when it is clear from context. Using these features, we can define the following social factors for the proposer agent, denoted $s_1, s_2, s_3$ that match the features we used to analyze human behavior in Section 3.

– Selfishness measures the extent to which proposers cared about their individual benefit.

$$s_1(x) = PO_P(x) - NN_P$$

[5] This phenomenon, deemed the "endowment effect", has been documented in the psychology literature [6]).

– Helpfulness measures the extent to which proposers were interested in the welfare of the group as a whole, as well as their own benefit.

$$s_2(x) = (PO_P(x) + PO_R) - (NN_P + NN_R)$$

– Competitiveness measures the extent to which proposers cared to do better than others. Such participants were willing to sacrifice some of their own benefit in order to increase this difference.

$$s_3(x) = (PO_P(x) - NN_P) - (PO_R(x) - NN_R)$$

For each potential exchange $x_j$, we defined a "social" utility function $u(x_j)$ for a general proposer player that is a weighted sum of the features defined above:

$$u(x_j) = \sum_{i=1}^{3} w_i \cdot s_i(x_j)$$

where $w_i$ denotes the weight associated with social factor $s_i$.

This utility function is transformed into a stochastic model that assigns a probability to each possible exchange at each round of interaction. A soft-max function is used to make the likelihood of each exchange proportional to the likelihood of other possible exchanges. This model is well suited for capturing certain aspects of human behavior: The stochasticity of the soft-max function allows for proposers to deviate from choosing the action associated with the highest utility, but in a controlled way. In addition, the likelihood of choosing an exchange that incurs a high social utility will increase if there are few other similar exchanges that incur high utility, and will decrease if there are many other similar exchanges.

The model parameters, represented by the feature weights $w_1, \ldots, w_3$ were trained using supervised learning. The labeled training set consisted of the exchanges made by proposers in the task and table conditions. Each instance consisted of pairs of possible exchanges $(x_*, x_j)$, where $x_*$ was the offer made by the proposer, and $x_j$ is any other possible exchange. To estimate the feature weights of the utility function, we used a gradient-descent technique that learned to predict the probability of a chosen offer $x_*$ given any other offer $x_j$ as follows:

$$P(x_* \text{ chosen } \mid x_* \text{ or } x_j \text{ chosen}, \mathbf{s}_*, \mathbf{s}_j) = \frac{1}{1 + e^{u(x_*) - u(x_l)}}$$

Here, $\mathbf{s}_*$ denotes the social factors associated with the offer that was proposed. This probability represents the likelihood of selecting $x_*$ in the training set, given $x_j$. The error function to minimize is defined as the extent to which the model is not a perfect predictor of this concept,

$$err_j = 1 - P(x_* \text{ chosen } \mid x_* \text{ or } x_j \text{ chosen } .\mathbf{s}_*, \mathbf{s}_j)$$

Taking the derivative of this function, we obtain the following update rule for the features $\mathbf{w}$, where $\alpha$ is a constant learning rate, and $\mathbf{d} = \mathbf{s}_* - \mathbf{s}_j$.

$$\mathbf{w} = \mathbf{w} + \alpha (err_j)^2 \cdot (1 - err_j) \cdot \mathbf{d}$$

We learned separate models for the task and table contexts. In both cases, we trained and tested the algorithms separately, using ten-fold cross validation. We obtained the following average posterior parameter values for the features selfishness, helpfulness and competitiveness in each condition.

| Condition | Learned weights |
|---|---|
| Task | $(5.20, 3.2, 0.40)$ |
| Table | $(8.20, 1.3, 8)$ |

As shown in the table, both task proposers and table proposers are selfish, in the sense that they place high weight on their own benefit. However, table proposers assign higher weight to their own benefit than do task proposers, suggesting they are more selfish than task proposers. Task proposers also assign a higher weight to helpfulness and significantly lower weight to competitiveness than table proposers. These values align with the trends reported in the Results and Analysis section.

We evaluated both models on test sets comprised of held out data from both task and table conditions. We report the average negative log likelihood for all models in the following table as computed using ten-fold cross validation. A lower value for this criteria means that the test set was given a higher likelihood by the model.

| Training / Testing Condition | Average Log Likelihood |
|---|---|
| Task / Task | 0.144 |
| Table / Task | 1.2 |
| Table / Table | 0.220 |
| Task / Table | 1.2 |

As shown by the table, the model trained and tested on the task condition was able to fit the data better than the model trained and tested in the table condition, indicating that computer agents participating in mixed human-computer task settings must model human performance in a way that reflects the context under which the decision was made.

In addition, the model trained in the task condition outperformed the model trained in a table context when both models were evaluated in task contexts. (And conversely for the model trained in the table condition.) The extent to which both models underperformed when evaluated in the context they were not trained on was similar for both conditions. These results clearly imply that the context in which decisions are placed affects the performance of computer models that learn to interact with people.

## 5 Related Work

A series of studies spawned by the seminal work of Tversky and Kahneman [11, 7] show that the way decisions, outcomes, and choices are described to people influence their behavior, and these different "framings" fundamentally affect people's perceptions and conceptualizations. For example, people's decision-making is sensitive to the presentation of outcomes as losses or wins and to the presence of alternative choices [12]. In addition, decisions are influenced by the labeling of interactions with terms that carry cultural or social associations [8]. Some of these framing effects (e.g., presence of alternatives) abstract away from domain specifics, while others (e.g., social associations) typically rely on real world or domain knowledge and experience, sometimes quite subtly. Both types of framing effect may be investigated using CT. For example, we have conducted a preliminary study of the effects of social relationships on decision-making in CT [9].

Our work is fundamentally different from work that addresses the effects of graphical versus tabular representations on people's decision-making [13, 5]. This work has shown that performance on particular tasks is enhanced when there is a good match between the mode used to represent a task and the cognitive resources required to complete it. It aims to present information in a way that provides good "cognitive fit", a vivid representation that overcomes the constraints of human information processing. In contrast, we examine whether the structural features that are inherent in task contexts, such as the relationship between goals and resources, affect people's decision-making. We do not address the cognitive-load implications of different contexts or with their mode of representation. In fact, we control for the effects of cognitive load in both task and table settings by providing participants with decision-support tools.

Lastly, recent work on modeling the social factors that affect people's decision-making behavior have concentrated on task contexts only [9, 3]. This work extends these approaches by comparing models of decision-making in task contexts and table contexts.

## 6 Conclusion and Future Work

We have shown that when making decisions placed in the context of a task setting, people behave more helpfully, less selfishly, and less competitively than when making decisions in the context of a table of payoffs. Further, people are significantly more likely to behave according to game theoretic equilibria in table contexts, which has a negative effect on their performance, compared to their behavior in task contexts. Moreover, people do not behave differently in task contexts when they are given access to the possible payoffs for themselves and others. We induced predictive models of the decision-making processes, showing that when learning in task contexts, computer players are better at predicting people's behavior than when learning in completely abstract contexts.

The results reported in this study suggest that when building a system for human-computer interaction, placing the decisions in task contexts will improve the performance of both people and computer agents that learn from people. Therefore, designers of systems that involve people and computers interacting together need to decide how to appropriately contextualize the decisions they present to participants.

While our experiments were performed in a relatively simple and flat task context, the fact that differences were found in this context suggest that it is likely there will be even greater ones in more complex settings. Our results provide a guideline for agent designers, specifically that the right context should be used when investigating human decision-making processes. We have presented an infrastructure for conducting such an investigation, and a methodology for how it might be done.

## 7 Acknowledgments

Development and dissemination of the Colored Trails formalism is supported in part by the National Science Foundation under Grant No. CNS-0453923 and IIS-0222892.

## References

1. Rajarshi Das, James E. Hanson, Jeffrey O. Kephart, and Gerald Tesauro. Agent-human interactions in the continuous double auction. In B. Nebel, editor, *Proc. 17th International Joint Conference on Artificial Intelligence (IJCAI'01)*, 2001.
2. Y. Gal and A. Pfeffer. Predicting people's bidding behavior in negotiation. In P. Stone and G. Weiss, editors, *Proc. 5th International Joint Conference on Multi-agent Systems (AAMAS'06)*, 2006.
3. Y. Gal, A. Pfeffer, F. Marzo, and B. Grosz. Learning social preferences in games. In *Proc. 19th National Conference on Artificial Intelligence (AAAI'04)*, 2004.
4. B. Grosz, S. Kraus, S. Talman, and B. Stossel. The influence of social dependencies on decision-making. Initial investigations with a new game. In *Proc. 3rd International Joint Conference on Multi-agent Systems (AAMAS'04)*, 2004.
5. S. L. Jarvenpaa. The effect of task demands and graphical format on information processing strategies. *Management Science*, 35(3):285–303, 1989.
6. D. Kahneman, J.L. Knetsch, and R.H. Thaler. The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5:193–206, 1991.
7. D. Kahneman and A. Tversky, editors. *Choices, Values, and Frames*. Cambridge University Press, 2000.
8. V. Liberman, S. Samuels, and L. Ross. The name of the game: Predictive power of reputation vs. situational labels in dertermining prisoners' dilemma game moves.
9. F. Marzo, Y. Gal, B. Grosz, and A. Pfeffer. Social preferences in relational contexts. In *Fourth Conference in Collective Intentionality*, 2005.
10. M.E. Pollack. Intelligent technology for an aging population: The use of AI to assist elders with cognitive impairment. *AI Magazine*, 26(9), 2006.
11. A. Tversky and D. Kahneman. The framing of decisions and the psychology of choice. *Science*, 211:452–458, 1981.
12. A. Tversky and I. Simonson.
13. I. Vessey. Cognitive fit: A theory-based analysis of the graphs versus tables literature. *Decision Sciences*, 22(2):219–240, 1991.