Modeling User Perception of Interaction Opportunities for Effective Teamwork

Ece Kamar, Ya'akov Gal and Barbara J. Grosz
School of Engineering and Applied Sciences
Harvard University, Cambridge, MA 02138 USA
{kamar, gal, grosz}@eecs.harvard.edu

***Abstract*—This paper presents a model of collaborative decision-making for groups that involve people and computer agents. The model distinguishes between actions relating to participants' commitment to the group and actions relating to their individual tasks, uses this distinction to decompose group decision making into smaller problems that can be solved efficiently. It allows computer agents to reason about the benefits of their actions on a collaboration and the ways in which human participants perceive these benefits. The model was tested in a setting in which computer agents need to decide whether to interrupt people to obtain potentially valuable information. Results show that the magnitude of the benefit of interruption to the collaboration is a major factor influencing the likelihood that people will accept interruption requests. They further establish that people's perceived type of their partners (whether humans or computers) significantly affected their perceptions of the usefulness of interruptions when the benefit of the interruption is not clear-cut. These results imply that system designers need to consider not only the possible benefits of interruptions to collaborative human-computer teams but also the way that such benefits are perceived by people.**

I. Introduction

This paper considers collaborative decision-making in heterogeneous groups of computer agents and people. Participants of a collaborative activity work together towards satisfying a joint commitment, but they also adopt and care about their individual goals. To be successful in such settings, agents need to reason about the effects of their actions on the collaboration and the way other participants perceive these actions. This paper proposes a new model for collaborative decision-making that significantly reduces the complexity of making joint decisions. The model distinguishes between those situations in which agents' actions affect the collaboration, and those in which their actions affect only their individual tasks. It decomposes the collaboration into smaller interacting subproblems that can be analyzed independently and combined to capture the collaborative utility of an action.

When people participate in collaborative activities with computer agents it is necessary for the agents to reason about the ways in which people perceive the utility of the collaboration and its constituent actions. We empirically investigate the mismatch between the actual utility of an action in a collaborative context and people's perception of it, exploring the different factors that may influence people's perception of this utility. The failure to consider this mismatch may cause a person to reject a valuable interaction opportunity, thereby turning what could have been a beneficial interaction for the collaboration into a performance degrading disturbance.

Our investigations use interruption management as an example of a decision making capability needed for collaborative activities in which agents are distributed, conditions may be rapidly changing and decisions are made under uncertainty. Interruptions are important for effective collaborative work, because agents often possess information required by others on their team. However, interruptions are inherently disruptive. If they are not managed and timed properly, they may negatively affect the emotional state and awareness of the user and may reduce the overall task performance of the user and the system [1].

For example, a writer's collaborative assistant that autonomously searches for bibliographical and citation information needs to know when to ask user for information and how to time requests [2]. If the assistant continuously asks whether to cite each paper that meets a user's keywords and commands it will disrupt the user's writing process.

Our decoupling model for collaborative decision making, called DECOP, synthesizes techniques from decision theory and computer science, but adapts them to collaborative contexts. DECOP takes into account the costs and benefits to all participants, whether human or computer agent, so that decisions to interrupt are based on the collaborative benefit to the group. Unlike previous models of interruption management, DECOP also reasons about the possible mismatch between a computer's estimate of the utility of an interruption and a person's perception of it. It focuses on determining the factors that influence people's perception of interruptions, and their tendency to accept or reject them when they are generated by a computer system.

We constructed a computer agent based on DECOP model in an empirical collaborative setting to investigate the way people perceive interruptions. We analyzed the effect of three factors on human perception of the usefulness of interruption requests: the magnitude of the interruption utility, the timing of interruptions and the perceived type of the partner (human or computer agent). The results revealed that the magnitude of the utility of interruption is the major factor affecting the likelihood that people will accept interruptions. The results also indicate that the perceived partner type and cost of the interruption to the subject affect people's perception of the usefulness interruptions when the utility gain is less clear-cut.

II. Related Work

Computing joint strategies for multi-agent decision-making problems under uncertainty has shown to be infeasible for realistic problems [3]. In this work, we introduce a new approach for efficient collaborative decision making in the spirit of near-decomposable models [4]. The remainder of this section compares our approach to the previous work in the domain of interruption management.

A key aspect of reasoning about interruptions in collaborative settings is the ability to accurately estimate the costs and benefits of the interruption to all parties so that the outcome of the interruption positively affects group task outcomes. Previous work on adjustable autonomy identifies the points at which it is most suitable to initiate interactions with a person, but does so without relating this decision to a user's mental state or the task being performed. Interruptions are driven solely by system needs and managed based on benefit to the system [5]. Prior work on interruption management has addressed user needs, but has focused mostly on the effect an interruption has on a person's cognitive state, rather than the benefit of interruptions to collaborative activities [6]. Few models have combined these two aspects into an integrated decision making mechanism [7], and none have done so in the kinds of rapidly changing domains of uncertainty we consider.

While there has been significant work on mixed-initiative system design, there has been little empirical work on how people perceive interruption utilities and make interruption decisions in human-computer interaction settings. Avrahami et al. [8] investigated the differences between a person's self report of interruptibility and other people's predictions about that person's interruptibility. However, this work considered face to face human interaction, rather than human-computer interaction. Gluck et al. [9] focused on designing notification methods to increase human perception of utility whereas Bunt et al. [10] showed that displaying system rationale to people may induce a person to trust a computer system more.

III. The Interruption Game

This section describes a game designed for investigating the interruption management problem in a setting that does not require sophisticated domain expertise. The "interruption game" involves two players, referred to as the "principal" and the "agent". Each player needs to complete an individual task but the two players' scores depend on each other's performance making this a collaborative endeavor.

The game is played on a board of 6x6 squares. Each player is allocated a starting position and a goal position on the board. The game comprises a fixed, known number of rounds. At each round, players advance on the board by moving to an adjacent square. The players' goals move stochastically on the board according to a Gaussian probability distribution centered at the current position of the player. [1] Players earn 10 points in the game each time they move to the square on which

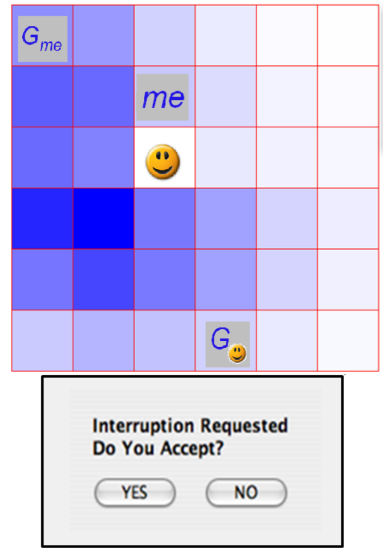

Fig. 1. Game Screen-shot: *me* is the principal player, *smiley* is the agent player, G_{me} is the principal's goal, G_{smiley} is the agent's goal. The degree to which each square is shaded represents the agent's uncertainty about its goal. Dark squares imply higher certainty.

their assigned goal is located, and the goals are reassigned to random positions on the board. Players can see their positions and the goal location of the other player, but they differ in their ability to see their own goal location: The principal can see the location of its goal throughout the game, while the agent can see the location of its goal at the onset of the game, but not during consecutive rounds.

At any round, the agent can choose to interrupt the principal and request the current location of its goal. The principal is free to accept or reject an interruption request. If the principal rejects the interruption request, the players continue moving. If the interruption is accepted by the principal player, the location of the agent's goal in the current round (but not in consecutive rounds) is automatically revealed to the agent. There is a joint cost for revealing this information to the agent, in that both participants will not be able to move for one round. The game scenarios used in the empirical evaluation are simplified to allow a single interruption through the game.

The rules of the game also require the agent to provide the principal with its belief about the location of its goal. This information may influence the principal's decision about whether to accept an interruption. A snapshot of the game from the perspective of the principal player is shown in Figure 1. The rules of the game provide incentives to players for reaching their goals as quickly as possible and interruptions initiated by the agent are critical determinants of players' performance. The agent's uncertainty about the location of its own goal increases over time, and its performance depends on successfully querying the principal and obtaining the correct position of its goal.

The game is collaborative in that the score for each player depends on a combination of its own performance and the performance of the other player. The players share a joint

[1] The movement of the goal is restricted in that it does not move closer to the position of the player.

score function that is the cumulative score of both players. An interruption is potentially beneficial for the individual performance of the agent, who can use this information to direct its movement, but it only induces the principal's performance negatively. Providing this information is costly for both players. When the agent deliberates about whether to ask for information, or when the principal deliberates about whether to reveal the information to the agent, the players need to weigh the trade-offs associated with the potential benefit to the agent player with the detriments to their individual performance in the game. The success of both players in the game depends on the agent's ability to estimate the collaborative value of interruption at each point in the game and use that information to choose when to interrupt the principal.

The interruption game is not meant to be a complete model of any specific domain or application. Its purpose is to provide a simple setting in which to study the factors that influence interruption management in collaborative settings. It provides a setting that is analogous to the types of interactions that occur in collaborative settings involving a mixed network of computer agents and people. For example, the principal player in the interruption game may represent the user of a collaborative system for writing an academic paper, and the agent may represent the collaborative assistant responsible for obtaining bibliographical data. While both of the participants share a common goal of completing a document, each of them must work independently to complete its individual task, such as composing paragraphs or searching for bibliographical information. This aspect is represented in the interruption game by assigning an individual goal for each player. The movement of these goals on the board corresponds to the dynamic nature of these tasks. For example, the user may not know what to write next, and the system may have uncertainty about search results. The agent's lack of information about its own goal location in the game corresponds to the uncertainty of a system about the preferences and intentions of its user, such as which bibliographical information to include in the paper. The ability to query the user for keywords and to choose among different bibliographies provides the system with valuable guidance and direction. It may, however, impede the performance of both participants on their individual tasks, because the system needs to suspend its search for bibliographical data when it queries the user, and the user may be distracted by the query. This dynamic cost of interruption represents the costs incurred to both users and computer agents due to task switching and task recovery for initiating and responding to an interruption.

IV. Modeling Interruption Opportunities

In this section we formalize the interruption game as a multi-agent decision-making problem under uncertainty, and we provide efficient methods for computing players' estimates of the benefit of interruptions in the game. These models are not meant to predict how people play the interruption game or respond to interruption requests in general. Rather, they provide a way to compute a theoretical baseline, which is a fully rational computational estimate for the value of interruption in the game. In the empirical section, we use this baseline to enable empirical analysis of human behavior in mixed-initiative settings. This analysis allows the study of the efficacy of these models when they are used by computers to interact with people in the game under various experimental conditions.

The interruption game can be modeled as a Decentralized Markov Decision Process (Dec-MDP) [11], a formalism for multi-agent planning that captures the collaborative nature of the interruption game and its associated uncertainty. A Dec-MDP includes a set of states with associated transition probabilities, a set of actions and observations for each agent, and a joint reward function. A solution of a Dec-MDP is an optimal joint policy for all agents that is represented as a mapping from states to actions.

To model the interruption game, the state space of the Dec-MDP will combine all of the information relating to the tasks of both players, including their positions on the board, the positions of their goals, the current round and the belief of the agent about its own goal position. The solution of the Dec-MDP assigns a policy to the agent that initiates interruption requests when they are expected to result in a benefit to both players according to the joint reward function and assigns a policy for the principal to accept interruption requests that have actual positive benefit. Unfortunately, finding optimal solutions to Dec-MDPs is NEXP-complete [3]. The size of the state space makes it infeasible to compute the complete joint policy for both players in the interruption game.

However, our goal is not to exhaustively compute optimal policies in the interruption game, but to be able to generate interruptions when they are perceived to be beneficial to the collaboration. We hypothesized that such interruptions would be likely to be accepted by people. Our novel model for decision making, called DECOP, exploits an important characteristic of this game and many domains in which computer agents and people work together: When players are not making or replying to interruption requests, they are performing their individual tasks, and each player needs to consider only its individual score in the game. In this case, the two tasks are essentially independent, and they can be solved separately. As the agent can only interrupt the principal once, the expected utility of an interruption can be computed efficiently, because an interruption request will render the two tasks independent from the interruption moment until the end of the game. At each turn, the policy for the agent is to interrupt and request information from the principal when it is deemed beneficial for both participants. The next section details DECOP model that captures the benefit of an interruption by solving the individual tasks for both participants in the game, and combining these solutions in order to devise strategies for interruption management in the game.

A. Computing a Policy for the Principal

The principal has complete information about the game, so its task can be modeled as a Markov Decision Process (MDP). Let B denote the set of board positions; $|B|$ denotes the size of the game board; $p \in B$, $g \in B$ are the positions of the principal and its goal respectively; $m \in A$ is a movement action of the player; $P(g', p, g)$ is the probability of the goal position moving from position g to position g' when the player is in position p. The state space of the MDP includes every possible position of the principal and its goal at each round. We denote $S_P^h = \langle p, g \rangle$ to be the state at round h and induce a state transition function T that assigns a probability to reaching state S_P^{h+1} from S_P^h given action m. T can be directly derived from P. The reward function R assigns the score in the game for reaching the goal if an action transitions a player to its goal square, and 0 otherwise.

Let Π_P^* denote the optimal policy for the principal player in the game. The value $V^{\Pi_P^*}(S^h)$ of this policy at state S_P^h maximizes the reward at state S_P^h for an action m and future states given the transition probability function,

$$V^{\Pi_P^*}(S_P^h) = max_m[R(S_P^h, m) + \sum_{S^{h+1}} T(S^{h+1}, m, S_P^h) \cdot V^{\Pi_P^*}(S_P^{h+1})] \quad (1)$$

We compute the optimal policy and its value using ExpectiMax search. In this process we grow a tree with two types of nodes, decision nodes and chance nodes. There is a decision node for each state in the MDP, and each child of a decision node is labeled with a movement action for the principal. Chance nodes represent moves of nature, and each child of a chance node is labeled with a possible movement of the goal of the principal, and is assigned a probability according to the transition function. When traversing the tree, we recursively compute a value for each chance node that is a weighted average of the value of each of its children according to its probability. We compute a value for each decision node by choosing the child with the maximal value, and select that action. With memoization, the number of nodes generated by the search is bounded by $|B|^2 \cdot |H|$, which is polynomial in the number of rounds in the game.

B. Computing a Policy for the Agent

The agent cannot observe the position of its goal on the board, and without interrupting the principal it receives no information relating to this position. We model its task as a No Observation Markov Decision Process (NOMDP), which is a special case of an MDP with no observations. The state space for this model includes the position l of the agent on the board, its belief $b \in \Delta B$ over its goal position, and current turn h. We denote the state for the agent as $S_A^h = \langle l, b \rangle$. As we are modeling the agent's individual task, rather than its interaction with the principal, we leave out the interruption action and use the set of actions A and reward function R identical to the ones described for the principal player. The agent updates its belief b to $b^{'}$ after each turn according to the goal movement distribution P as follows:

$$\forall c' \in B, b'(c') = \sum_{c \in B} b(c) \cdot P(c', l, c) \quad (2)$$

The value of an optimal policy for the agent Π_A^* at state $S_A^h = \langle l, b \rangle$ can be computed using Equation 1, substituting Π_A^* for Π_P^* and S_A for S_P. Because the belief of the agent about its goal position is incorporated into the state space, there are an infinite number of states to consider, and using ExpecitMax in a straightforward fashion is not possible. However, applying the belief update function after each turn, only a small number of states turn out to be reachable. The deterministic belief update function maps each combination of states with full information (i.e., states in which the agent knows the correct position of its goal) and the number of turns since full information to a single belief state, thus to a single state. As a result, we can grow the search tree "on the fly", and only expand those states that are reachable after each turn. Memoization is not possible in this technique, and thus the complexity of the complete search is exponential in the length of the horizon.

C. Computing the Benefit of Interruption

To compute the benefit of an interruption, its effect on both the agent's and the principal's individual performance must be taken into account. It is the aggregate of these two effects that determines the utility of interruption. An interruption is initiated by the agent, but it is only established if the principal accepts it. The effect of an interruption on the individual game play of a player is the difference of the values of two states; one in which an interruption is established, and other in which it is not. Given the principal and its goal are located on squares p and g respectively in game round h, let $EU_P^{NI}(S_P^h = \langle p, g \rangle)$ denote the expected utility of the principal when it is not interrupted, and pursues its individual task. This is equal to the value for the principal of carrying out its individual task as shown in Equation 1. Thus we have,

$$EU_P^{NI}(S_P^h) = V^{\Pi_P^*}(S_P^h) \quad (3)$$

For the agent that does not get to observe its own goal position, let $EU_A^{NI}(S_A^h = \langle l, b \rangle)$ denote the expected utility of the agent when it is not interrupted, and pursues its individual task. This is the value to the agent of carrying out its individual task:

$$EU_A^{NI}(S_A^h) = V^{\Pi_A^*}(S_A^h) \quad (4)$$

Let $EU_P^I(S_P^h = \langle p, g \rangle)$ denote the expected utility for the principal when it accepts an interruption. If the principal player is interrupted, it cannot move for one round, but its goal may move stochastically according to the probability distribution P. We denote the new goal position as g^{h+1}. Given our constraint that there can only be one interruption made in the game, the benefit of interruption for the principal is the expected value of its individual task in future rounds, for any possible position of its goal. Formally, the utility of interruption for state S_P^h, denoted $EU_P^I(S_P^h)$ is computed as follows:

$$EU_P^I(S_P^h) = \sum_{g^{h+1} \in B} P(g^{h+1}, p, g) \cdot V^{\Pi_P^*}(S_P^{h+1} = \langle p, g^{h+1} \rangle) \quad (5)$$

If the agent successfully interrupts the principal, the principal will reveal the position of the agent's goal. The agent will update its belief about its goal position in the following round, and use this belief to perform its individual task in future rounds. However, when it deliberates about whether to interrupt in the current round, it needs to to sum over every possible position of its unobserved goal, according to its belief about the goal location. Let $S_A^h = \langle l, b \rangle$ be the current state of the agent, including its position on the board and belief about its goal position. Let g denote the current position of its goal. The expected value for interruption for the agent is denoted EU_A^I and is computed as follows:

$$EU_A^I(S_A^h) = \sum_{g \in B} b(g) \cdot V^{\Pi_A^*}(S_A^{h+1} = \langle l, b' \rangle) \qquad (6)$$

Here, b' refers to the belief state of the agent in which probability 1 is given to g, the true position of its goal as revealed by the principal, and updated once to reflect the stochastic movement of the goal in turn h.

D. Deciding Whether to Interrupt

By combining the expected values of the principal and agent players with and without interruption, it is now possible to compute the agent's estimate of the benefit of an interruption. We denote $EBI_P(S_P^h)$ to be the expected benefit of interruption for the principal given S_P^h, which is simply the difference in utility of the principal between accepting and interruption and carrying out its individual task.

$$EBI_P(S_P^h) = EU_P^I(S_P^h) - EU_P^{NI}(S_P^h) \qquad (7)$$

The expected benefit of interruption for the agent is denoted $EBI_A(S_A^h)$ and is computed similarly:

$$EBI_A(S_A^h) = EU_A^I(S_A^h) - EU_A^{NI}(S_A^h) \qquad (8)$$

The interruption game is collaborative in that the combined performance of both participants determines their individual scores. The agent can observe the state S_P^h of the principal, and for any combined state $S^h = (S_P^h, S_A^h)$, the agent will consider the joint expected benefit of interruption to both participants, EBI, and choose to interrupt if this joint benefit is positive.

$$EBI(S^h) = EBI_P(S_P^h) + EBI_A(S_A^h) \qquad (9)$$

The agent cannot observe the correct position of its goal and estimates the benefit of interrupting under this uncertainty. Thus, not every interruption initiated by the agent is truly beneficial for the team. In contrast, the principal can observe the position of agent's goal and can capture the actual benefit of the interruption, denoted ABI with certainty. Any interruption with positive ABI offers a positive expected benefit to the team. The value of ABI is the sum of the individual benefits of interruption to both the principal and the agent.

Let g_a be the agent's goal position, the actual benefit of interruption for both participants given states S_P^h and S_A^h is

$$ABI(S^h) = EBI_P(S_P^h) + EBI_{P,A}(S_A^h) \qquad (10)$$

Here, the term $EBI_{P,A}(S_A^h)$ refers to the principal's perception of the agent's benefit from revealing the goal position g_a, where l is the current position of the agent, b' refers to the belief state of the agent in which probability 1 is given to g_a and updated once.

$$EBI_{P,A}(S_A^h) = V^{\Pi_A^*}(S_A^{h+1} = \langle l, b' \rangle) - EU_A^{NI}(S_A^h) \qquad (11)$$

The advantage of DECOP model introduced above is that it reduces the complexity of the multi-agent decision making process to that of two separate single agent decision making processes. Because the agent is allowed to interrupt only once during a game-play, the decoupling method is able to accurately capture the benefit of an interruption initiated by the agent. DECOP model assumes that principal players are fully rational and computationally unbounded. In the case of such players, we would expect any interruption with positive ABI to be accepted, and any interruption with negative ABI to be rejected. However, people may not be fully rational or computationally unbounded, and we expect people's perception of the benefit of interruptions to differ from baseline values calculated by DECOP model. In the empirical investigations described in the next section, these baseline values are compared with the subject responses to detect the possible mismatch between a computer's estimate of the benefit of an interruption and a person's perception of it, and to identify a subset of factors that affect the way that humans perceive the effectiveness of interruptions.

V. Empirical Methodology

This section uses strategies derived from DECOP model for playing the interruption game to explore the way people make decisions in an empirical setting. A total of 26 subjects participated in the study. The subjects were between ages of 19 and 46 and were given a 20 minute tutorial of the game. Subjects played between 25 to 35 games each, and were compensated in a manner that was proportional to their total performance.

During the empirical evaluation, all subjects were allocated to play the roles of principals, while the role of the agent was assigned to a computer that used the methodology described in the previous section to play the interruption game. Each game proceeded in the manner described in Section III. In particular, the agent could not observe its own goal location, but is allowed to initiate an interruption once to acquire the correct location of its goal from the principal. At each turn of the game, the policy of the agent is to interrupt only if the expected benefit of an interruption (EBI) is positive

Interruptions were generated by the computer agents at different points in the game with varying actual benefits, game levels and perceived partner types. We measured people's responses to these requests given the game conditions at the time of interruptions, which included the number of turns left to play, the positions of both players on the board, and the agent's belief about the location of its goal.

A principal player that uses DECOP model to determine whether to accept an interruption request is perfectly rational in that is uses the collaborative benefit of interruption (ABI), given in Equation 10, as the sole factor for this decision. However, we expected people to differ from this rational model. The purpose of the empirical work is to measure the extent to which different factors in the game, such as the collaborative benefit of interruption (ABI), the timing of interruptions and the perceived partner type, influence people's perception of interruptions. To investigate the way subject responses change with respect to benefit of interruptions, the game scenarios were varied to have different ABI values. To investigate the effect of the timing of an interruption on the subjects' likelihood of acceptance, we varied the level of the game that an interruption is initiated. Lastly, we expected that the type of agent player (whether a computer or human) would affect the way people respond to interruption requests. For this reason, subjects were told they would be interacting with a human for some games, however they were always paired with an agent [2].

Subjects were given randomly generated game scenarios that vary the actual benefit of interruption to both participants (ABI) to cover four types of values: -1.5 (small loss), 1.0 (small gain), 3.5 (medium gain), 6.0 (large gain). These values represent the smallest and largest benefit values that can be generated from interruptions with positive expected benefit (EBI), which is a necessary condition to initiate interruption requests by the agent player. The levels in the game in which interruptions occurred varied to cover the beginning of a game (level 3), the middle of a game (level 5) and the end of a game (level 7). There were 540 game instances played when the perceived agent was a computer (PP:Computer) and 228 data points when the perceived agent was a person (PP:Person).

VI. Results and Discussion

The following results analyze a total of 768 game instances collected in our study. Table 1 shows interruption acceptance rates for different levels and ABI values for the same game instances when perceived partner type (PP) is person or agent. The optimal policy for the principal player is to accept an interruption if its associated benefit (ABI) is positive and to reject otherwise.

PP:Computer	Level 3	Level 5	Level 7
ABI -1.5	0.16	0.16	0.41
ABI 1.0	0.27	0.7	0.81
ABI 3.5	0.91	0.97	0.79
ABI 6.0	0.91	0.95	0.95
PP:Person	**Level 3**	**Level 5**	**Level 7**
ABI -1.5	0	0.11	0.44
ABI 1.0	0.72	0.94	1
ABI 3.5	0.91	0.94	0.88
ABI 6.0	1	0.88	1

TABLE I
Acceptance Rate of Interruptions

[2]Approval was obtained for the use of human subjects in research for this misinformation.

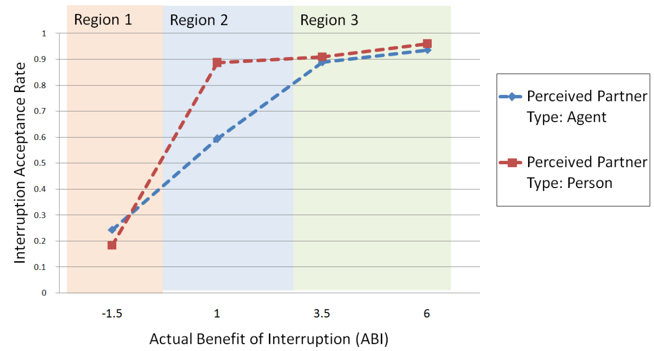

Fig. 2. Effect of interruption benefit and perceived partner type on interruption acceptance rate

As the results of Table 1 show, the utility of an interruption is the major factor influencing the probability that an interruption will be accepted by a person. The interruption acceptance rate increases significantly as the benefit of interruption rises from -1.5 to 1.0 ($p < e^{-20}$, α=0.001) and from 1.0 to 3.5 (p=0.0013, α=0.01). However the rise from 3.5 to 6.0 does not further improve the acceptance rate. These results confirm that people were successful at perceiving interruption benefits above a certain threshold. Similarly, when an interruption is costly for the collaboration, people are significantly more likely to reject the interruption. However, subjects varied in their responses to interruptions offering slightly positive gains, indicating the difficulty to estimate the benefit of interruption when its usefulness is ambiguous.

Figure 2 summarizes the acceptance rates of interruption as a function of the actual benefit of interruption and perceived partner type (person vs. computer). We divide the figure into three regions of interruption benefits: small losses (Region 1), small gains (Region 2), and large gains (Region 3). The analysis shows that for large losses (Region 1) and for small gains (Region 3), changing the perceived partner type does not affect the likelihood that the interruption will be accepted. In contrast, for interruptions offering small gains (Region 2), the acceptance rate is significantly larger if the perceived partner type is a person ($p = 3 \times 10^{-5}$, $\alpha = 0.001$). This result implies that when the benefit of interruption is straightforward, people do not care whether the initiator of the interruption is a person or a computer. However, for those cases in which the benefit of interruption is ambiguous, people are more likely to accept interruptions that originate from other people. This result suggests that agent designers need to consider the way they present interruptions to their users in cases where the perceived benefit is ambiguous. It also aligns with recent studies showing that mutual cooperation is more difficult to achieve in human-computer settings as compared to settings involving people exclusively [12].

Figure 3 shows the effect of interruption timing (the level of the game) on people's acceptance rates for interruptions of small losses and small gains (The interruption timing does not affect the acceptance rate for interruptions of large gains). We

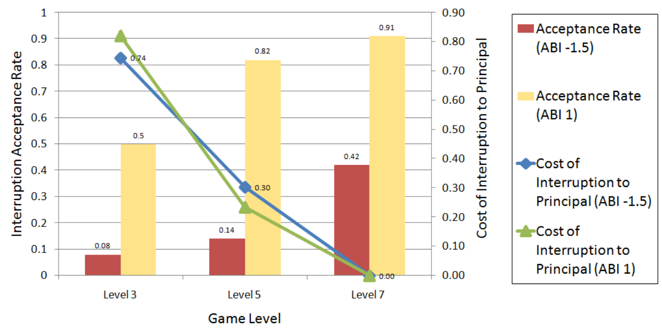

Fig. 3. Correlation of interruption acceptance rate with the cost of interruption to subjects (-ABI_P) for small gains and losses.

expected that interruptions occurring late in the game (i.e., with fewer number of turns left in the game) will be more likely to be accepted when they incur positive benefit, and rejected when incurring a loss. However, as shown by the Figure 3, as the game level increases, so does the acceptance rate for interruptions of both small losses (ABI -1) and small gains (ABI 1). There is a significant increase in the acceptance rate when game level increases from 3 to 5 (p=0.002, α=0.01) and from 5 to 7 (p< 10^{-6}, α=0.001).

One factor that may explain the correlation between the acceptance rate and the game level for interruptions of small gains and losses, is the cost of interruption to the subject. As shown in Figure 3, the cost of interruption to the subject (ABI_P) decreases as game level increases. Thus, for interruptions of small gains and losses, we found that the acceptance rate is negatively correlated with the cost of interruption to the principal. In addition, it was revealed that the benefit of the interruption to the principal (ABI_P) is a better predictor of the acceptance rate than ABI_A, the benefit of interruption to the agent (logistic regression $SE = 0.05, R^2 = 0.19, p < 0.001$). Thus, human subjects tend to overestimate their own benefit from interruptions as compared to the benefit for the agent. Consequently, the benefit of interruption to person may be weighted more in person decision making model than the benefit of the interruption, and people may be more likely to accept an interruption with low ABI_P among interruptions with identical benefit. Further study is required to determine whether these conjectures hold and better understand the correlation of acceptance rate with the cost of interruption to the person. This conjecture is supported by some subject responses to survey questions regarding their strategies for accepting interruptions. Answers include:

"If the agent was in the totally wrong direction and I had several moves left, I would allow the interruption. I always wanted the sure thing for myself."

"If the collaborator was way off in knowing and had enough moves to likely catch it after I told the location, I accepted. If it compromised my ability to get my goal, I declined."

Lastly, we emphasize that these results are a first step in understanding the human perception of interruptions in collaborative settings. Our goal was not to design computational strategies directly applicable for interruption management in real world domains, but rather to show that effective interruption management needs to consider the collaborative benefit of interruption to both user and system, and to point system designers to the types of factors that people consider when they reason about interruptions. In future work, we plan to extend the study to better understand the effects of computational and cognitive complexity on people's interruption strategies, focusing on the possible role of trust and overestimation of costs.

VII. Conclusion

We have presented a novel model for collaborative decision-making that computes the utility of actions from the possibly different perspective of the participants of the collaborative activity. We evaluated the model in a specially designed human-computer collaborative setting in which computers need to manage their interruption requests to humans. We showed that the actual benefit of interruptions to both computer agents and people is the major factor affecting the likelihood that people will accept interruption requests. However, for those cases in which the benefit of interruption is ambiguous, people prefer to accept those interruptions that originate from other people.

Our empirical studies provide a number of insights about human decision making in the context of human-computer collaboration in dynamic, uncertain domains. Understanding the way that people make interruption decisions will enable the development of better mechanisms for initiating interruptions, focusing on the interruptions that are more likely to be accepted. These investigations will provide the foundation for building agents that collaborate with people efficiently without overburdening them.

References

[1] P. Adamczyk and B. Bailey, "If not now, when?: the effects of interruption at different moments within task execution," *CHI*, 2004.

[2] T. Babaian, B. Grosz, and S. Shieber, "A writer's collaborative assistant," in *Proceedings of the 7th international conference on Intelligent user interfaces*. ACM New York, NY, USA, 2002, pp. 7–14.

[3] D. S. Bernstein, S. Zilberstein, and N. Immerman, "The complexity of decentralized control of markov decision processes," in *UAI*, 2000.

[4] H. Simon, "The architecture of complexity," *Proceedings of the American Philosophical Society*, pp. 467–482, 1962.

[5] P. Scerri, D. Pynadath, W. Johnson, P. Rosenbloom, M. Si, N. Schurr, and M. Tambe, "A prototype infrastructure for distributed robot-agent-person teams." in *AAMAS*, 2003.

[6] E. Horvitz and J. Apacible, "Learning and reasoning about interruption," *ICMI*, 2003.

[7] M. Fleming and R. Cohen, "A user modeling approach to determining system initiative in mixed-initiative ai systems," in *UM*, 2001, pp. 54–63.

[8] D. Avrahami, J. Fogarty, and S. E. Hudson, "Biases in human estimation of interruptibility: effects and implications for practice," in *CHI*, 2007.

[9] J. Gluck, A. Bunt, and J. McGrenere, "Matching attentional draw with utility in interruption," in *CHI*, 2007.

[10] A. Bunt, J. McGrenere, and C. Conati, "Understanding the utility of rationale in a mixed-initiative system for GUI customization," in *UM*, 2007.

[11] E. Kamar and B. J. Grosz, "Applying mdp approaches for estimating outcome of interaction in collaborative human-computer settings," *MSDM*, 2007.

[12] J. Rilling, A. Sanfey, J. Aronson, L. Nystrom, and J. Cohen, "neural correlates of theory of mind within interpersonal interactions," pp. 1694–1703.