Oh no you didn't! Mediating Conflict Between Humans and Intelligent Virtual Agents

Sajika Gallege
1210 West Dayton Street
Madison, WI 53715
+1 608 262-1204
sgallege@cs.wisc.edu

Eric Lederer
1210 West Dayton Street
Madison, WI 53715
+1 608 262-1204
elederer@cs.wisc.edu

ABSTRACT
As embodied agents, both physical and virtual, representing intelligent systems become more self-empowered in their collaboration with users, there rises an increased need to manage emergent communicative breakdowns due to conflicts of interest and belief. We examined interactions between participants and a virtual agent and whether classic conflict mediation strategies such as relationship-affirmation, dominance, and rationalization could be successfully employed by the virtual agent to promote collaborative solutions to decision-making conflicts within the experimental task. Our results demonstrate that relationship-affirming strategies emphasizing teamwork and the integrity of the relationship produced positive effects in encouraging cooperative work.

Keywords
Conflict, conflict mediation, conflict mitigation, conflict resolution, virtual agent, human-computer interaction.

INTRODUCTION
Conflict rises inevitably out of group interaction. [1, 12] The emergence of conflict can stem from fundamental differences in values ("what should be"), beliefs ("what is"), or control over common resources, but it's often simply enough that the members of the group differ in individual preferences and taste. [2] Given autonomy of action, thought, or both, we observe conflicts as a natural consequence. So what happens to human-computer interaction when the computer system is deemed intelligent enough to act and think for itself?

In our work, we consider the emergence of conflict in a human-computer cooperative task with an onscreen virtual agent. Embodied agents, both virtual and physical (think robots), enable a powerful communication channel between humans and intelligent computer systems beyond traditional user interfaces.

However, as these agents become ever more self-empowered and sophisticated, there rises an increased need to manage emergent communicative breakdowns due to conflicts of interest and belief with human users.

While it is conceptually trivial to incite conflict, it is imperative from a design perspective that we develop effective strategies to handle it as it arises. Organizational research focusing on conflict resolution identifies and addresses two key issues involved in the successful mediation of conflict:

- How do we prefer to approach and handle conflict, *as individuals*? [3]
- What strategies can we employ to mitigate conflicts? [4]

We apply existing conflict resolution theory to our problem involving the virtual agent. First, we distinguish individual participants by their preexisting predilection for certain styles of approaching conflict using a validated instrument, the Thomas-Kilmann Conflict Model. [3] Second, we empower the virtual agent to employ classic conflict mediation strategies such as relationship-affirmation, dominance, and rationalization to promote a collaborative solution to a decision-making conflict within the experimental setting.

Finally, we explore the success of these strategies at resolving the decision-making conflict. Do users respond to these inherently social mediation techniques when employed by a semi-social virtual agent? If so, does the efficacy of a particular resolution strategy vary across users by their preferred methods for handling conflict? We hope that by addressing these questions, we will be able to help inform better design of embodied agents representing intelligent computer systems.

RELATED WORK
Similar research in managing agent breakdowns in agents has been conducted by Lee and others with Carnegie Mellon's Snackbot. [6] Rather than focusing on conflicts of interest and belief between the system and its user, the researchers instead address the fallibility of the novel

technology and the negative impact that breakdowns in service have on user perceptions of the agent. They demonstrate that pre-breakdown *expectancy-setting* strategies (such as warning user's of the robot's limitations) and post-breakdown *recovery* strategies (such as apology and compensation) can successfully mitigate user dissatisfaction. They also demonstrate that the efficacy of individual strategy varies across users by personal characteristics. [6]

Beyond HCI, the field of organizational management offers a broad body of research on task-oriented group conflict resolution; Jones' and White's 1985 study on conflict resolution styles and task effectiveness demonstrated that those with a high need to defer authority and decision-making responded positively to forcing (akin to our agent's dominant strategies) and that those with a high need for affiliation responded positively to smoothing (akin to our agent's relationship-affirming strategies). [9] Combining these findings with Reeves' and Nass' *media equation*, which demonstrated that users respond to social cues and behaviors from technological artifacts, [8] we expected the agent's conversational script and appearance to evoke social responses to its use of human-like mitigation strategies, and that these responses would extend the same behaviors observed in conflict scenarios in organizational management involving only human interaction.

METHODOLOGY

Experimental Design

Our experiment explores the roles of gender, an individual's preferred approach to conflict resolution, and each of three different conflict mitigation strategies in affecting user's propensity to collaborate with the agent, as measured by the user's responsiveness to the agent's attempts at resolving a decision-making conflict. We conducted a mixed-model experiment: within-participants in that each participant was exposed to all 3 mitigation strategies, between-participants in that we compared differences across participant by gender and preferred approach to conflict resolution, as determined by the Thomas-Kilmann Conflict Model.

Factor I: Preferred methods of handling conflict

The Thomas-Kilmann Conflict Model (TKCM) represents a dual-concerns model mapping a space of "desire to satisfy one's own concerns" against "desire to satisfy the other party's concerns" in a conflict scenario. A validated instrument exists to map users to the model space, thereby suggesting their preferred approaches to conflict resolution by five distinct categories. [3, 5, 10, 11] By completing the post-survey, we scored users towards five points within the model space (Figure 1):

- Competitive users (higher interest in satisfying one's own goals, goal-oriented)
- Avoidant users (recognizing the existence of conflict, but preferring to shy from resolution)
- Accommodative users (higher interest in satisfying the other party's goals, relationship-oriented)
- Compromising (recognizing the existence of conflict, preferring to compromise whereby both parties make some sacrifice)
- Collaborative (recognizing the existence of conflict, preferring to search for resolutions fully satisfying both parties)

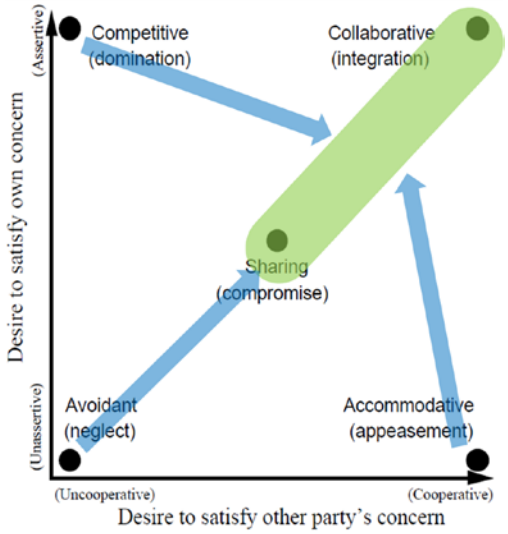

Figure 1. The Thomas-Kilmann Conflict Model. (Figure courtesy of Easterbrook. [12])

Towards each of the five points, users are scored over a scale of 0 and 12, where higher values predict greater preference for and likelihood to employ that particular style of handling conflict. The instrument designates the highest-ranked style as the "dominant," but adds that a "support hierarchy" exists whereby users fall on secondary preferences on a situational basis. [3] In our analysis, we consider a users' preference for a particular style using this same scale of 0-12, discretizing users into low (0-6), high (7-12) ranges. For example, a user with a value of 10 towards Accommodation and 3 towards Compromise was designated as Accommodation-high and Compromise-low. This technique provided for finer grain analysis of users with complex competing style preferences.

While user TKCM scores did not impact eligibility for participation, we focused our analysis on users outside the most integrative space of the model space—the highly competitive, avoidant and accommodative users—with our motivating intent being to drive these users towards collaborative behaviors supporting both parties' interests.

Factor 2: Conflict mitigation strategy

Throughout the course of the experiment, each participant was exposed to three classic conflict mitigation strategies; preexisting work in humans led us to form hypotheses about the expected outcomes of employing these strategies during the decision-making conflict.

I. Relationship-affirmation: These strategies emphasize the importance of the relationship between the two parties involved in the conflict. For example, one of the agent's actual appeals follows:

   "I'm sorry, I know it's a burden and you'd rather not be scrubbing floors, but I'm not excited about that job either. Our best chance at throwing a successful party is to work together as a team. If you take care of scrubbing floors, then I could handle making the shopping list. Is that ok?"

II. Rational Forewarning: These strategies emphasize knowledge sharing; the agent provides logical and statistical information about the task and emphasizes the agent's specialized knowledge of the domain. For example, one of the agent's actual appeals:

   "Based on my previous work with human partners, I have determined that I handle scrubbing floors at a rate 41% faster than my human partners. On the other hand, it takes me 27% longer than the average human partner to finish making the shopping list. I suggest that we switch these to finish sooner overall, but the final decision is up to you."

III. Agent-assertive strategies/dominance: These strategies urge the user to comply with a decision made by the agent towards a solution deemed more desirable by the agent, emphasizing the agent's specialized knowledge of the domain. Unlike rational forewarning, the agent does not provide information to inform user decision-making, but requests the user to defer decision-making. For example, one of the agent's dominant appeals:

   "So making the shopping list might seem easy, but it's actually a very big responsibility and can have a big impact on whether or not we throw a successful party. There's more to it than you think. Just let me handle making the shopping list and you can focus on scrubbing floors. I leave the decision ultimately up to you though."

Factor 3: Gender

Given the demonstrated importance of gender differences in human-computer interaction research, we consider gender of participants as well to capture any effect it may have on responsiveness to the agent's mediations.

Hypotheses

We framed our hypotheses with respect to outcomes of the individual mitigation strategies.

I Relationship-affirmation: We expected to find that the agent's use of such behaviors would have a positive effect in moving accommodative users towards the integrative space of the model. Studies in conflict resolution demonstrate that users who prefer accommodative handling of conflict resolution tend to make sacrifices of their own goals so as to avoid threats to the relationship, [12] suggesting that relationship-affirming mitigation strategies might be successful at allaying these fears and improving collaboration.

II Rational Forewarning: Given the foreignness of the experimental setting in which users were collaborating with the agent, we expected rationalization to produce positive effects across TKCM types by providing information that would be perceived as useful.

III Dominance: We expected dominance to produce positive effects in moving avoidant users towards more collaborative solutions by obviating the need for the user to make decisions [11] or to deal with the emergent conflict between parties.

Experimental Procedure

Participants were enlisted using Amazon Mechanical Turk [7] upon condition that they be rated 98% or higher within the Mechanical Turk system and that they be logged in from a location in North America. After consenting to participate, they were presented with two screens of instructions introducing the task, the preference ranking protocol (details to follow), and the virtual agent. The participants were then given three training modules to familiarize themselves with the controls of the game and to allow them to begin to form impressions of tasks they might prefer over others.

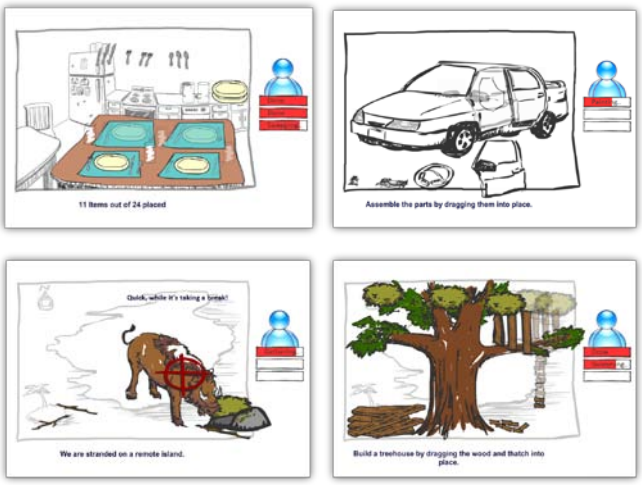

Figure 2. Actual interfaces of 4 (out of 18 possible) subtasks completed by participants.

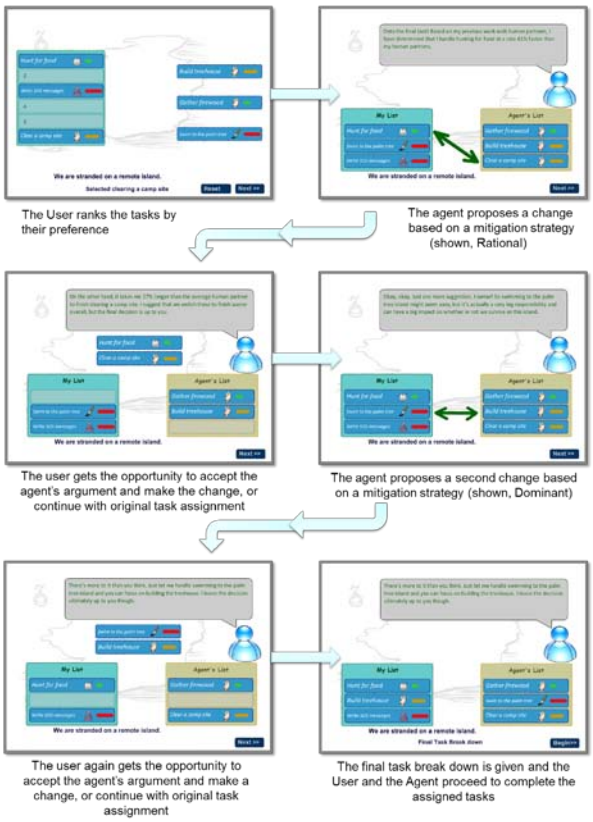

Figure 3. Task selection interface in the "Island Survival" scenario

Upon beginning the game, users were presented with one of three randomly-selected scenarios (assembling a car, throwing a house party, or surviving on an island) comprising six subtasks to be completed: 3 by the user, 3 by the agent. Users were told to rank by *preference* the six subtasks, after which the system presented the six subtasks as two separate lists, those subtasks to be completed by the user and those to be completed by the agent. A conflict was introduced, suggesting that the agent had different values, preferences, or beliefs about the domain and how the tasks should be assigned. The agent then randomly proposed, from the three above mentioned mitigation strategies, a solution in the form of a reassignment of tasks which the agent believed which effect a better outcome in the interests of both parties. The users were given the opportunity to accept or reject the agent's argument. The users were then exposed to a *second* randomly-selected mitigation strategy and proposal reassigning two subtasks, repeating the above procedure. Following this second round of mediation, the users set to completing their three tasks, while progress bars in the interface indicated the agent's progress towards completing its 3 assigned tasks.

The tasks were of four general categories, each of which was introduced to the user during the training: drag and drop, text entry, simple click, and multi-click/paint. Tasks varied dramatically in difficulty and length of time to complete so as both to encourage users to form preferences and to instill in users a sense that decisions made during the task breakdown were not inconsequential.

If the user completed his or her tasks before the agent (as indicated graphically by the progress bars), he or she was forced to wait while the agent completed its work. Otherwise, the user proceeded to the second (and finally, third) of the scenarios, each time following the task breakdown procedure described above.

Participants were exposed to a total of six attempts at conflict mediation by the agent: two in each of three total scenarios. They were presented in random order with two relationship-affirming arguments, two dominance arguments, and two rational forewarning arguments. During the task breakdowns, subtasks were initially presented in random order, and the script was designed to provide enough flexibility that any individual argument in any scenario could appear sensible to the participants.

After completing the third scenario, users proceeded to complete a post-survey designed both to assess their perceptions of the activity and the agent, as well as to deploy the 30-question Thomas-Killman Conflict Instrument (TKCI) to assess their preferred method of approaching conflict.

Mean completion time of the experimental task itself was around 15 minutes (M = 903 sec, SD = 233.2 sec). Mean completion time for the post-surveys including TKCI instrument was around 8-9 minutes (M = 510.52 sec, SD = 318.84 sec). The full sample consisted of 23 males and 41 females. All participants were paid $3.00 for their time.

Measurements

Objective

For each of the agent's six attempts at conflict mediation, we measure the user's response as a single objective dependent variable: the decision made by the user to either a) cooperate by reassigning the tasks as the agent proposes, or, b) reject the agent's proposal by restoring the subtasks to their original user-ranked assignment.

Subjective

We included 16 questions on 7 point Likert scales to evaluate each participant's perception of the tasks and the agent. The task-related questions measured the importance of task difficulty, type [single click, drag and drop, text entry, multi-click/paint], the real-world nature of the task, and agent suggestions regarding task assignment. The agent-related questions included social and behavioral characteristics such as friendliness, helpfulness, aggressiveness, rationality and intelligence.

RESULTS

Objective

We conducted a mixed-model analysis of variance (ANOVA) to test how the mitigation strategy (within-participants factor) and accommodativeness (between-participants factor) affected participants' acceptance of agent arguments. (Figure 4(a)) The analysis showed that, overall, participants displayed significantly higher acceptance of arguments when the agent used relationship-affirming strategies as opposed to rational mitigation strategies, (F[1,126] = 4.24, p < .05) or dominant mitigation strategies (F[1,126] = 4.06, p < .05).

The analysis also showed that *highly-accommodative* participants showed significantly higher acceptance of arguments using relationship-affirming strategies than when rational mitigation strategies, (F[1,126] = 7.94, p < .01). Post hoc pairwise tests showed that dominant mitigation strategies were more effective on highly-accommodative users than dominant mitigation strategies p=.09. (Figure 4(b)) The above finding supports our first hypothesis that the relationship-affirmation would be effective on highly-accommodative users.

We conducted another mixed-model analysis of variance (ANOVA) to test how mitigation strategy (within-participants factor) and gender (between-participants factor) affected participants' acceptance of agent arguments. (Figure 4(c)) The analysis showed that female participants were significantly more receptive to arguments using relationship-affirming strategies than to ones using rational mitigation strategies, (F[1,126] = 5.34, p < .05). There were no significant main or interaction effects on male participants.

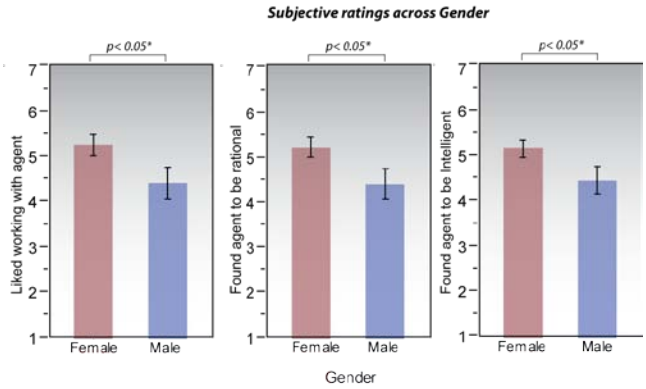

Figure 5. Subjectively-rated perceptions of the agent across gender.

Subjective

We analyzed the subjective responses between-participants by gender and preferred conflict handling styles. The analysis showed that female participants rated enjoyment working with the agent significantly higher than did male participants (F[1,63] = 4.38, p < .05). (Figure 5) The analysis also revealed that the female participants found the agent significantly more intelligent than did male participants (F[1,63] = 4.40, p < .05) and that female participants found the agent significantly more rational than did male participants (F[1,63] = 4.85, p = .05).

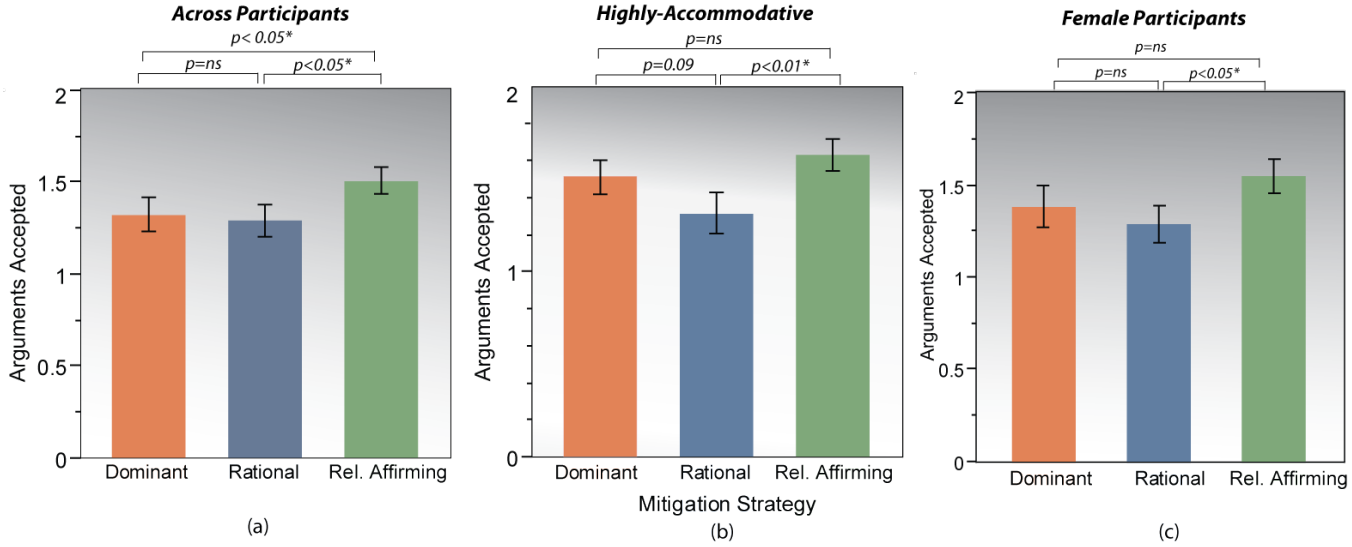

Figure 4. (a) Acceptance rate of arguments by mitigation strategy across all participants. (b) Acceptance rate of arguments by mitigation strategy within highly-accommodative users. (c) Acceptance rate of arguments by mitigation strategy within female participants.

The analysis of subjective responses revealed that the highly-avoidant participants found the agent significantly more rational than did participants with lower avoidant scores (F[1,63] = 5.29, p < .05), as well as significantly more selfish (F[1,63] = 4.08, p < .05). (Figure 6)

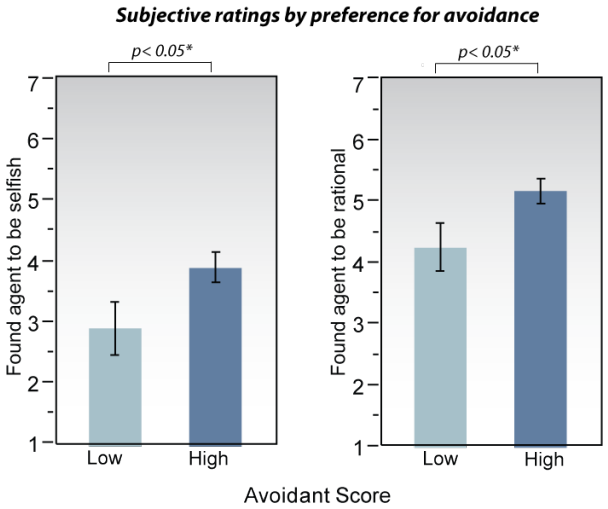

Figure 6. Subjectively-rated perceptions of the agent by preference for avoidance (low- or high-).

DISCUSSION

Of our three hypotheses, our results showed strong support for our first hypothesis alone. The results of the ANOVA (Figure 4(b)) document a statistically significant difference within highly accommodative participants with respect to their receptiveness to relationship-affirming mediations over the rational forewarning mediations which we tested with the agent (F[1,126] = 7.94, p < .01). Finer-grained analysis of these findings (taking gender into account) did not suggest that this result was simply due to the modalities of the sample distribution. We believe that this attests to statements made in our hypothesis: that accommodative users are easier to appeal to by emphasizing the importance of the relationship with the agent. Here, we remark on several other factors that might contribute to this result. First, ANOVA involving all users (Figure 4(a)) demonstrated the relationship-based mediations to prove more successful overall than the other two strategies. We have not to date empirically tested whether the relationship-affirming scripts are more *inherently* convincing than those used for dominance or forewarning, or whether they are more convincing appeals *as employed by the (or for that matter, an) agent.* Second, we concede that the experiment would benefit from a cleaner distinction made between the introduction of the conflict itself and the agent's act of mediation. Some of what we observe may measure the user's willingness to comply with the agent, rather than an eagerness to truly collaborate, particularly relevant in drawing conclusions about the behavior of accommodative or obliging individuals. We discuss this more in our recommendations for further work.

We did not find evidence to support our second hypothesis. While we expected participants to respond positively to statistically-grounded arguments because humans expect computers to analyze and to express information numerically, users were relatively unmoved by rational appeals. This may be due to failure of the script to convince as discussed above. It may also represent distrust of the value of statistical knowledge held by an agent embodied as a rather simplistic cartoon, inhabiting a playful cartoon environment. It may also be that in executing non-technical tasks such as setting a kitchen table, users find the presentation of statistical information unrelatable, inappropriate, or as one participant described, "insincere."

We also did not find evidence to support our third hypothesis that dominance would produce favorable results in highly-avoidant users. Much of our belief that dominant-assertive mediations would succeed with these users was founded upon the idea that this style of mediation would free users from the perceived need to handle the conflict by effectively removing the call for decision-making. However, in standardizing our experimental conditions with the other two mediations, our procedure still required the user to *accept* the agent's rise to dominance, with full knowledge that the final decision ultimately remained with the user. We continue to believe that a dominant mediation can be successfully employed by an agent if properly implemented.

Though not related to one of the stated hypotheses, we did discover an interesting difference among male and female participants. As mentioned, overall sample receptiveness to relationship-affirming strategies appeared to be significantly higher than to rational forewarning strategies, (F[1,126] = 4.24, p < .05), to dominant mitigation strategies, (F[1,126] = 4.06, p < .05). However, further stratifying these results across gender (Figure 4(c)) revealed that this applied with statistical significance only in female participants, where acceptance of relationship-affirming arguments was significantly higher than of rational forewarning arguments, (F[1,126] = 5.34, p < .05). We interpret this in conjunction with significant findings of our post-survey scale ratings of perceptions of the agent (Figure 5), in which female participants also rate the agent significantly higher in terms of its intelligence, (F[1,63] = 4.40, p < .05), and their enjoyment in working with it, (F[1,63] = 4.38, p < .05). We believe that despite the semi-social behavior and appearance of the agent, female participants were more likely to make sense of their collaboration with the agent as a social interaction, which seems critical to one's willingness to accept a relationship-affirming appeal. Nonetheless, that they found the agent to be more rational, (F[1,63] = 4.85, p = .05), than did male

participants did not improve their receptiveness to the agent's rational mediations.

FURTHER WORK

A refinement of the decision-making setup should consider further distinguishing the emergence of the conflict itself—the revelation that the agent's values and beliefs differ from those of the user—with the actual act of mitigation in the form of the agent's intervention. In its current state, the experiment may not make it clear enough that the agent's argument for a change to the task breakdown represents its attempt at mediation, rather than its request of compliance or deference from the user. Refining this distinction should allow us to more conclusively claim that responsiveness to the agent's suggestions indicates a desire to collaborate.

Despite assurances of the validity of the Thomas-Kilmann Conflict Instrument found in the literature, it was our experience that the distribution of users across types skewed heavily towards Compromising and Avoidant users, with very few participants registering primarily as Competing and Accommodative types. It may be worth future researchers' time to consider alternative models of conflict handling; Daly nicely summarizes variations of the dual-concerns model along slightly different dimensions in his Doctoral thesis work. [11]

We recommend the addition of a control treatment to Factor 2, the agent's set of mitigation strategies. Having a baseline for analysis would further demonstrate the success of the experimental treatments. This would require a rebalancing of the treatments and a revision of the decision-making process.

In future work, we also recommend deeper exploration of the role of the agent's appearance with respect to its conflict mediation style, such that incongruities of style, script, and appearance do not negatively impact user perceptions of the agent's trustworthiness.

ACKNOWLEDGMENTS

We thank Professor Bilge Mutlu for his guidance throughout the course of the experiment, as well as the Fall 2010 CS-838 class at the University of Wisconsin-Madison for their suggestions and critique.

REFERENCES

1. Robbins, S. P. *Organizational Behaviour: Concepts, Controversies, and Applications*, (fourth edition). Prentice Hall, NJ, 1989.
2. Deutsch, M. *The Resolution of Conflict*. Yale University Press, New Haven, 1973.
3. Thomas, K. Conflict and Conflict Management, in Dunnette (ed.), *Handbook of Industrial and Organizational Psychology*. Chicago: Rand McNally, 1976.
4. Strauss, A. *Negotiations: Varieties, Contexts, Processes and Social Order*. Jossey-Bass Publishers, San Francisco, CA, 1978.
5. Easterbrook, Steve. Handling Conflict Between Domain Descriptions With Computer-Supported Negotiation. *Knowledge Acquisition: An International Journal*, Vol 3, pp. 255-289, 1991.
6. Lee, M.K., Kiesler S., Forlizzi J., Srinivasa S., and Rybski P. E. Gracefully Mitigating Breakdowns in Robotic Services, in *Proceedings of HRI*, 2010.
7. Amazon mTurk, https://www.mturk.com.
8. Reeves, B., & Nass, C. *The media equation: How people treat computers, television, and new media like real people and places.* New York: Cambridge University Press, 1996.
9. Jones, Robert E. and Charles S. White. Relationships Among Personality, Conflict Resolution Styles, and Task Effectiveness. *Group & Organization Management* 1985 10: 152, 1985.
10. Daly, Timothy Michael. *Conflict management in consumer behavior: Examining the effects of preferred conflict management style on propensity to bargain.* Thesis work. University of Western Australia, 2009.
12. Easterbrook, S. M., E. E. Beck, J. S. Goodlet, L. Plowman, M. Sharples and C. C. Wood. *A Survey of Empirical Studies of Conflict. in CSCW: Cooperation or Conflict?* London: Springer-Verlag, pp. 1-68, 1993.