Saying YES! The Cross-cultural Complexities of Favors and Trust in Human-Agent Negotiation

Johnathan Mell
University of Southern California
Los Angeles, CA, USA
mell@ict.usc.edu

Gale Lucas, Jonathan Gratch
Institute for Creative Technologies
Los Angeles, CA, USA
{lucas, gratch} @ict.usc.edu

Avi Rosenfeld
Jerusalem College of Technology
Jerusalem, Israel
rosenfa@jct.ac.il

***Abstract*—Negotiation between virtual agents and humans is a complex field that requires designers of systems to be aware not only of the efficient solutions to a given game, but also the mechanisms by which humans create value over multiple negotiations. One way of considering the agent's impact beyond a single negotiation session is by considering the use of external "ledgers" across multiple sessions. We present results that describe the effects of favor exchange on negotiation outcomes, fairness, and trust for two distinct cross-cultural populations, and illustrate the ramifications of their similarities and differences on virtual agent design.**

Keywords—affective computing; intelligent systems; cultural differences

I. INTRODUCTION

When humans negotiate, more is at stake than simply the outcome of the current deal. Each interaction alters the relationship that develops between each of the parties, changing the potential actions and expectations that they hold about future negotiations. In business literature, this ongoing interactive process broadly falls under the moniker of "relationship building", but there is a more fundamental importance to far-sightedness. Negotiators who look blindly at the current deal will miss out on integrative potential that exists only over multiple games—what is known as Pareto-efficiency-over-time [6].

Consider the example of two roommates who wish to split up some recent purchases. One roommate may prefer oranges, while the other prefers apples—it is therefore reasonable for the apples and oranges to be split "unfairly" such that the first roommate gets all of the oranges, but no apples. That way, both parties have gotten what they wanted, and "grow the pie" in such a way that they are both the happier for it.

But consider a slightly different scenario. Now, our roommates are attempting to split the use of a necklace. The fair solution would be each party to take half of the necklace. Obviously, this would be less than ideal if one wished to wear the item. Perhaps, however, roommate Alice wants the necklace for her date tonight. It would be reasonable for our other roommate Beth to allow Alice full use of the necklace now, because Beth expects she will be given the same benefit later. Beth is creating a "ledger" in which she informally stores this as a "favor" that she is due later. Therefore, both parties have achieved Pareto-efficiency-over-time, even though their individual splits at a given point were very lopsided.

While this knowledge of favor-sharing has existed for some time in the business and negotiation literatures, applying these techniques to virtual agents is a relatively new development. When virtual agents are involved in the negotiation, there are further considerations, such as the importance of calculating the value of a ledger at a given time, and calculating Pareto-optimality-over-time for a given set of repeating negotiations. More fundamentally, humans negotiate with agents in subtly different ways than they negotiate with other humans, and affect and trust considerations require careful understanding and design.

Previous work [18] showed that a very simple text-based agent could alter the language used in order to successfully induce human players to discover Pareto-efficient-over-time solutions. However, understanding the long-term effects of this behavior on trust-building is a critical component to making "thoughtful" relationship-based agents. Culture-specific differences, such as propensity to be "promotion-focused" (hopeful, and risk-seeking) or "prevention-focused" (fearful, and risk-avoiding), can further refine this picture of trust [12]. Therefore, a robust virtual agent design should be aware of those constants across negotiators, and be ready to adapt to culture-specific differences.

We build on these existing ideas by conducting a cross-cultural study in which we compare the characteristics of participants from the United States and from India—two populations that differ markedly in promotion/prevention focus. We contribute by showing that Indian participants show a marked increase in offer acceptance rates in our negotiation task compared to their American counterparts. Furthermore, we show that measures of generosity and trust differ as well. Additionally, we show that trust is tied critically to the congruence between word and deed—following through on favors builds trust, but lose either component and trust falls. Finally, we demonstrate that while Indian participants may be more likely to distrust offers from a betraying agent, they are also more willing to accept deals in order to avoid a loss.

II. BACKGROUND

A. Human-Agent Negotiation Games

Negotiation, both between two humans as well as between humans and agents, is an important research topic across multiple scientific domains. Multi-issue bargaining, in particular, serves as a de facto standard for research into social cognition, distributive problem solving and interpersonal skill-development [26].

This work was supported by the National Science Foundation under grant ISS-1211064 and the U.S. Army. Any opinion, content or information presented does not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

978-1-4799-9953-8/15/$31.00 ©2015 IEEE

Most automated approaches to negotiation have focused on agent-agent interaction and make strong limits on the type of information that can be exchanged between parties. More recently, there is growing interest in algorithms that can negotiate with people, either to resolve [15] or mediate conflicts [1], or to teach negotiation skills [6]. These agents incorporate more complex forms of signaling, such as emotional reactions to offers [4] or natural-language dialog [24]. Most of this multi-turn negotiation research, however, has focused on short-term interactions and single negotiations. The present work both expands the unit of analysis to multiple negotiations, and attempts to untangle the relationship between trust, culture, and risk avoidance behaviors.

Our immediate motivation behind this work is to inform the design of virtual human partners for teaching negotiation skills by validating the effectiveness of human negotiation tactics in a virtual context. While it is true that humans interact with agents differently from how they interact with other humans, evidence suggests that interacting with "virtual humans" results in more similar behavior than simply interacting with an agent that displays no human-like features. The authors can point to several benefits of having an agent as a negotiation partner in educational settings, not least of which is that agents are infinitely patient and always available (professional negotiation trainers are rarely either).

Beyond the differences between human and agent partners, we are interested in how tactics that are emotionally reliant (such as trust-building and favor exchange) are affected by cultural difference. Some research has already explored the potential of virtual humans for negotiation training [1,2], and this study aims to build upon this body of research and extend its applicability to situations where negotiations repeat over time. The dynamics of relationships between negotiating partners play a critical role in the final outcome of many multi-issue tasks; indeed the relationship between truthfulness and trust—as well as eventual outcome—has been well established in the business and negotiation literature [13].

Multi-issue, repeated negotiations have a key feature in that there may exist solutions over multiple games that allow both parties to "grow the pie" and receive better solutions. Our agent operates in this space with a human participant. The question of what may encourage these superior instances of "integrative potential" to be discovered by one or both negotiators is not fully understood. Discovering Pareto efficient solutions over time as well as Pareto efficient solutions occurring within a single negotiation have been examined previously [18], but the connection between trust and risk avoidance has not yet been fully examined.

The notion of signaling intention facilitates the favors and ledgers technique, in which one party may accept an unfavorable agreement in the current negotiation with the expectation that it will receive a similar treatment from its partner in a subsequent negotiation [20]. If favors are issued during negotiations that have little utility for the offering agent and are received during negotiations that have more utility, integrative potential can be achieved by both parties. However, such practices rely on trust, and violating the expectations established by prior signaling could be considered a betrayal.

In negotiations, participants may be classified as more "prevention-focused" or more "promotion-focused". Prevention-focused negotiators tend to avoid risky outcomes in order to avoid a loss, while promotion-focused negotiators are more risk-seeking—looking to secure a win [12]. Clearly, the interaction between the emotional antecedents of these behaviors (fear and hope) and levels of trust in an agent can be very significant.

B. Cross-cultural Features

Understanding the differences in negotiation styles across cultures is necessitates considering the features that define each culture. While such generalizations do not define the individual, they are helpful in designing agents that can adapt to their audience in a much more helpful manner. Indian negotiators, for example, have been observed to be more prevention-focused than they are promotion-focused. In other words, when entering into potentially risky situations, they become very loss-averse. Indians "play to not lose" rather than "play to win" [7]. For our agent, this might imply that when faced with ultimatum games, Indian participants would be less willing to make poor offers for fear of the offers being rejected. Negotiators that are prevention-focused will be very unwilling to make offers that they fear may be rejected, especially when a rejection would force both players to take their BATNA or Best Alternative To Negotiated Agreement (as in ultimatum games). However, the extent to which trust in the agent may mediate this behavior is not fully understood.

Other work on cross-cultural negotiation by Gunia [10] indicates that Indian negotiators may be slow to trust other people, preferring instead to trust institutions. Gunia makes the link between trust and joint gains (Pareto Efficiency), such that a lack of trust may lead to a lack of discovery of joint gains. To the best of the authors' knowledge, this work represents the first study examining whether these cross-cultural differences exist between human negotiators and agent counterparts.

C. Trust

Trust is a very context-specific term. However, within negotiations, trust often refers to individual trust, which is a measure of how much a negotiator expects the other negotiator to obstruct or assist with the first negotiator's goals. In this way, trust can be seen as a consideration in the traditional risk-reward conflict. Negotiating partners that are more trusted are often seen as being low-risk. For risky bargaining strategies, like favors-and-ledgers (which relies on the assumption that favors will eventually be repaid), this trust factor is of course critically important to cultivate.

However, trust can also refer to a more broad type of *institutional* or *organizational* trust, which includes concerns about the effectiveness of systems [16]. In this way, a person may be very willing to trust the efficiency of the legal system, but harbor some doubts about a particular attorney. Such kinds of trust can be about a more specific field, such as trust in automation, which has been measured in other work [19].

Even beyond generalized trust in automation however, there is some evidence that the trust in the system's creator may bleed over into trust for the agent itself due to the institutional or organizational factors [16,23]. If the virtual

agent is seen as being designed by a reputable institution, then the agent itself may (at least at first) be seen as a trustworthy *individual*, even if its actions would speak otherwise. As this effect may be largely unavoidable in virtual agent design, it does blur the line between institutional and individual trust. Furthermore, agents may be viewed as more trustworthy to begin with in situations where the average human would be seen as untrustworthy, a fact which emphasizes the need to examine trust with regards to human-agent interaction [17].

III. Experimental Design

A. Predictions

Given the conflicting factors of general propensity to trust across cultures, as well as the varied prevention and promotion focuses of different populations, the authors felt it critical to extend our original study on American populations to a cross-cultural study which included Indians. Virtual agents that perform well under different cultures are possible, but it is essential that they adapt to specific needs that must be identified experimentally. Furthermore, we are interested to see what artifacts the use of agents over human partners may introduce that differentiate our results from prior work.

The previous work by Gunia indicated that due to the collectivist nature of Indian culture, Indians may be less willing to trust individuals than Americans. We anticipate this same result in our subject pool:

Hypothesis 1: Indians trust their partners less than Americans.

However, our usage of repeated ultimatum-games presents a problem for Indian participants. Although they may be less willing to trust their partner, they are also prevention-focused, which means that they may be willing to accept offers in order to avoid angering them and losing out in later rounds. We believe that this effect may override, causing "hyper-generosity", in which Indian participants accept deals that are lower even than their BATNAs. This tendency to "just say YES" in order to preserve a relationship is a common theme in popular business advice, and may be even more powerful in repeated game contexts. In other words:

Hypothesis 2: Indians accept offers more than Americans, even if the offer is lower than their BATNA.

The lack of trust in the good intentions of their partner may eventually lead to poorer outcomes overall. Gunia indicates that a lack of trust leads to less joint value discovered in integrative situations. We believe that additionally, the compulsion to "say YES" may lead participants to resolve negotiations quickly, missing out on potential joint value. This leads us to our third hypothesis:

Hypothesis 3: Indians find less integrative potential/joint value than Americans.

The decisions and implications for trust are not made in a vacuum. The actions of the agent are important in maintaining or breaking individual trust. Because of this, we expect that agents that match word and deed by promising favors and then returning them will be rated highest on measures of trust and fairness. Deviation from this in either the language used or the quality of offers will result in a drop in trust. That means that an agent designed with these principles in mind will glean the largest amount of trust and fairness, which we state as Hypothesis 4:

Hypothesis 4: Favor language and good offers (favor-seeking agent) will result in the most trust and perceived fairness.

B. Game Design

To realize the multi-issue, multiple-negotiation domain that we explore, we used the Colored Trails testing framework. Colored Trails is a negotiation testbed for analyzing the strategies of participants, and has been used in several types of games, including revelation games [21,8]. Our design involved a version of the interface that was deployable via the web and customized to allow our agent to engage in multi-issue bargaining games.

In Colored Trails, players both start with a set amount of different-colored "chips". By expending a chip, a player can move one space on the board of a similar color, with the intent to move toward a goal location. In our version, the closer a player gets to the goal, the more points they receive (Figure 1). Players can propose and respond to deals in which they propose exchanges of chips. In this way, the chips represent multiple issues with several "levels". With two red chips, for example, there are three levels (0, 1, and 2 chips for the player).

Our design featured a set of 5 repeated games, in which the player acted as the respondent to 4 ultimatum games. If they accepted the offer, chips would be exchanged and both the player and agent would move, and then score points. If they instead rejected the offer, no chips were exchanged and both players were forced to move using only the chips they had at the start of the game. This result represented the BATNA. The player was not allowed to engage in cross-talk with the agent; rather, they were forced to respond to the offer, which is standard for ultimatum games. The agent, however, sent text-based messages at various points in the game, notably when it proposed offers and when it received the player's response.

This design means that the domain is characterized by multiple, single-round, multi-issue ultimatum games with varying BATNAs and varying payoffs. This is not the same as single, multi-round games, where negotiators often progress from one issue to the next. Critically, it allows for joint value to be discovered across games, and Pareto-efficiency-over-time to exist.

Fig. 1. Example game setup within the Colored Trails testing framework

C. Agent Design

To determine the implications of agent design on trust and integrative potential discovery, a set of simple text-based agents were designed that were capable of playing a set of ultimatum-style games online with human participants. The agents were designed to exhibit behavior based on a 2x2 factorial design. Favor-seeking and Cooperative agents both alternated good and poor offers, thus returning favors to the player and allowing them to find integrative potential. Betraying and Competitive agents never returned favors and always made offers that were well below the player's BATNA. The agents were simple in their behavior: each agent followed an unchanging schedule over the 5 games, offering agent-favoring deals on the first and third games, and offering either agent-favoring or player-favoring deals in the second and fourth games (depending on agent type).

Our other experimental dimension was the type of language used by the agent. Both the Favor-seeking and Betraying agents used favor-specific language that included phrases like "This goal is important to me. I hope you can accept this deal as a favor to me. I'll really owe you one." The Cooperative and Competitive agents only used more generic language, such as "I think this deal is acceptable." The language used to deliver each offer to the player was identical across subjects, but varied slightly from game to game. All agents also used the same language to properly acknowledge player actions. In this way, the agents were dependent slightly upon player choice, as a player acceptance would result in positive acknowledgment ("That's great!"), while a rejection would result in the opposite. These agents are summarized in Table I.

D. Experimental Design & Subject Pool Considerations

We divided our subjects by nationality, encompassing data from 269 Americans (151 male, 118 female), which were collected in an earlier study, and 146 Indians (106 male, 40 female). This allowed us to complete our 2x2x2 factorial experimental design, using offer quality, presence of favor language, and nationality as our independent condition variables.

TABLE I. AGENT TYPES

	Favors returned	Favors never returned
Favor framing	Favor-seeking	Betraying
No favor framing	Cooperative	Competitive

Whenever experiments are conducted using different subject pools, especially cross-cultural ones, it is important to verify the integrity of the pool if possible. Given that these subject pools were also comprised of online participants, it was felt that these considerations were particularly integral to address. When creating our version of the Colored Trails interface for web use, it was first tested internally using researchers present at our institution but who were not affiliated with the project. In additional to being integral for bug identification, this beta test also allowed us to verify that behavior of in-person subjects was similar to online subjects. We then used Amazon's Mechanical Turk service to recruit online subjects for both studies.

Several steps were taken to ensure participants from both countries were equally and properly motivated to perform to the best of their ability. All subjects in both the American and Indian Mechanical Turk studies were required to answer a set of "attention check" questions following the game. These questions verified that subjects were actively engaged in the game and not simply "clicking through" by asking such things as "How many games did you play?" Those subjects that did not successfully answer these questions were excluded from analysis. We found that the exclusions rate for each nationality were extremely similar (24.7% for Indians, 24.1% for Americans) indicating that both worker pools understood and participated in the task similarly.

We also adjusted the monetary incentives for our participants appropriately based on the culture being considered. Mechanical Turk workers were paid a flat rate for their participation. Due to the cost of living differences between American and India, this rate was adjusted to be 37.5% lower for Indians than for Americans, to avoid over-incentivizing participants to rapidly complete the task. To further incentivize participants to play seriously, we entered all subjects into a lottery. They were informed that their performance (the amount of points they received) would affect their chances of winning one of 10 additional $10 monetary rewards. This bonus dollar number was chosen as being more than 2x the flat participation rate for Americans.

While it is of course natural to assume that certain differences between Mechanical Turk workers and in-person recruits may exist (such as computer literacy, for one), we feel that these precautions eliminate many external concerns. Future studies would do well to be mindful of the specifics of their subject pool (and indeed, future directions of our work include comparing Mechanical Turk and locally-recruited populations).

IV. EXPERIMENTAL RESULTS

Our first result is a surprising one. While it was generally expected that US participants would be more trusting of their agent partner than Indians, the opposite result was actually discovered, $F(1, 330) = 10.26$, $p = .001$, $d = -0.384$ (Figure 2).

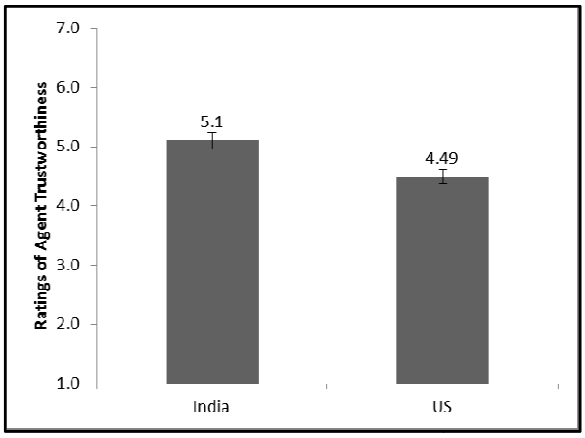

Fig. 2. Agents are seen as more trustworthy by Indians (7-point Likert)

This result will cause us to reject H1, which was based on earlier human-human studies. Our agent instead inspired trust in what is traditionally seen as a low-individual-trust culture.

To get a better view of the initial pattern of acceptances for various offers, we examine the result of game 3. We choose game 3 because it is the first point after which players may experience betrayal, which allows us to examine H2 for each agent type. We specifically look at players that have accepted an offer in game 1, then had that favor returned or betrayed (depending on condition) in game 2. We confirm that there is a main effect of nationality that shows that Indians accept offers more readily than Americans, χ^2 [1, N = 177], = 10.33, p = .001. This information is shown in Figures 3 and 4.

The difference between acceptance rates for favor-seeking and cooperative agents is signficant for Americans, as was previously reported. However, the result is only marginally significant for Indian populations, χ^2 [1, N = 177], = 3.13, p < .08. Figure 4 shows the difference between Competitive and Betraying agent acceptance rates. Although the previously reported result for Americans shows that there is a significant drop in acceptance rates due to the Betrayal agent, Indian participants show no significant difference. Indians appear to be relatively insensitive to betrayal, accepting even poor offers lower than their BATNAs readily, as predicted by H2.

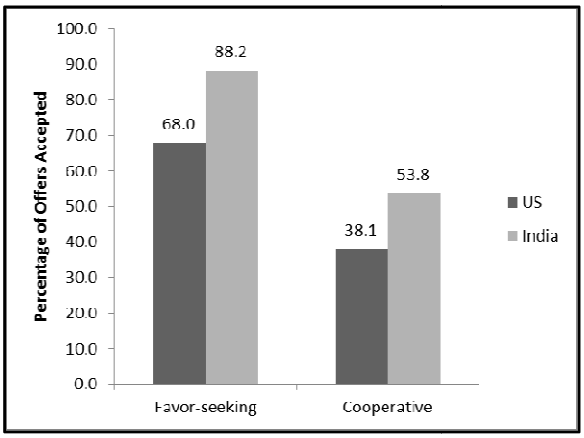

Fig. 3. Game 3: Main affect of nationality on acceptance rates is shown

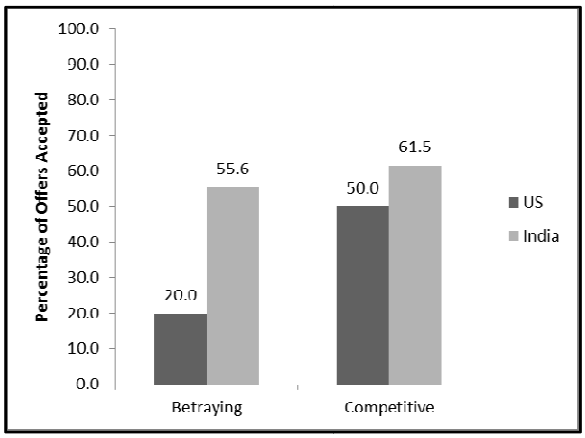

Fig. 4. Game 3: Indian participants are much less sensative to betrayal

We further examine the previous two sets of results by observing the relationship between trust, offer acceptance rates, and nationality. To verify that the increased trust by Indians is not spurious, we show significant correlation between trust and acceptance rates (game 3): F(1, 176) = 23.59, p < .001, d = 0.730. This result, combined with the main effect of nationality on acceptance rates and the effect of nationality on trust (F(1,176) = 44.88, p < .001, d = .642), leads us to explore mediation. To establish mediation, nationality would need to predict trust, and when entered simultaneously, trust should predict acceptance rates while the effect of nationality on acceptance rate is significantly reduced. Logistic regression predicting acceptance rates while controlling for trust showed a signficant effect of trust: B = .399, *Wald*(1) = 13.749, p < .001. It does not show full mediation, as the trust by acceptance rate interation remained significant: B = .717, Wald(1) = 4.51 p = .034, yet could indicate potential partial mediation. We confirm this using Preacher and Hayes' [22] SPSS macro for testing mediation using bootstrapping techniques [25], which revealed that the total indirect effect of trust was significant; the 95% BCa (bias corrected and accelerated) bootstrap confidence intervals were 0.176 to 0.742 (based on 1000 resamples). Therefore, we can conclude that acceptance rates are partially mediated by trust.

To address H3, the final round must be examined, in which the player crafted an offer as the proposer in the ultimatum game. The joint value found between the agent and the player represents the joint gain or integrative potential discovered. With no integrative potential discovered, both players would have to move only using their initial chips, each scoring 8 points (for a total of 16). Here, it was found that Americans found significantly more integrative potential than Indians, F(1, 331) = 22.26, p < .001, d = 0.533 (Figure 5).

Finally, to make the link between trust and agent behavior, measures of trust per agent type are displayed in Figure 6. Per hypothesis H4, trust is in fact higher for the favor-seeking agent for both Indians and Americans. We examined the favor-seeking agent condition using a 3v1 follow-up contrast to a one-way ANOVA—for Americans: t(1, 94) = -7.77, p < .001, and for Indians: t(1, 74) = -3.81, p < .001.

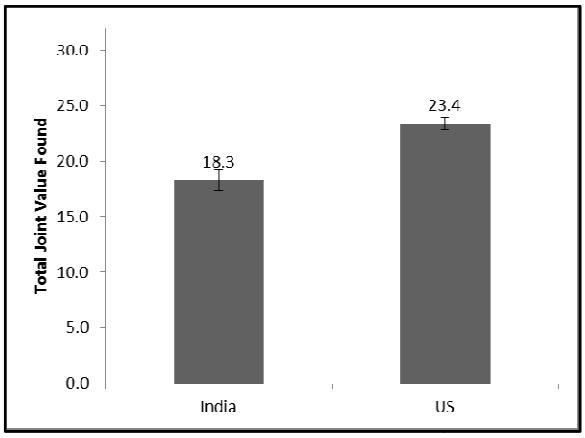

Fig. 5. Game 5: Total joint value found is higher among US players

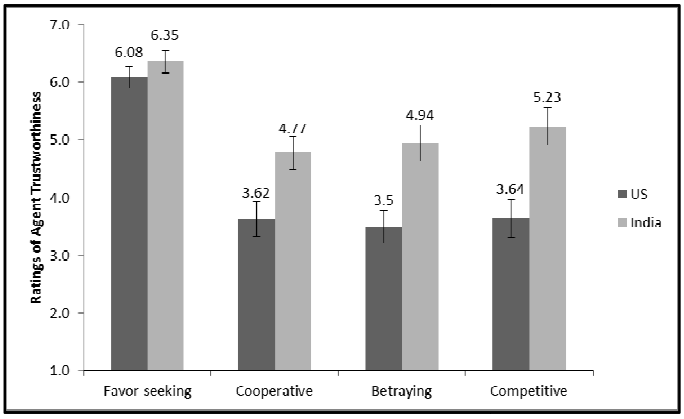

Fig. 6. Trust remains highest for favor-seeking agents in both cultures

These results were mirrored for measures of fairness, as displayed in Figure 7. Follow-up contrasts again showed that for Americans: $t(1, 94) = -7.73$, $p < .001$, and for Indians: $t(1, 74) = -3.67$, $p < .001$. No main or controlling effect of gender was found in any reported results.

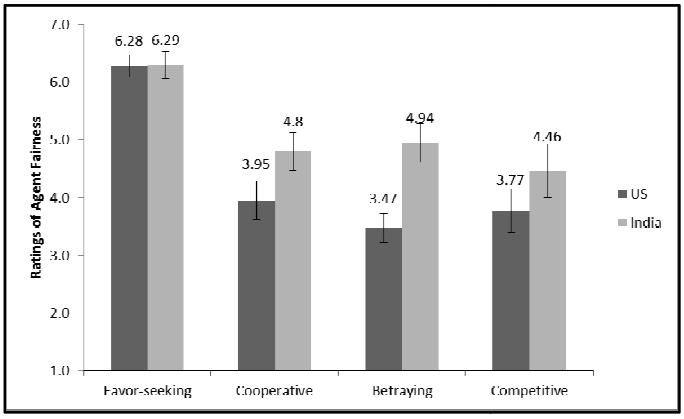

Fig. 7. Fairness remains highest for favor-seeking agents in both cultures

V. Discussion and Future Work

Although Indian subjects show several key differences compared to their American counterparts, several fundamental patterns remain very similar across cultures—a promising result for virtual agent design. Indians may be more willing to concede than Americans, but the long-term negative effects of "bad behavior" on relationships is very similar.

We found that Indians actually trust *more* than Americans, invalidating our earlier H1. This hypothesis was based on Gunia's human vs. human result, whereas this study examined human vs. agent negotiations across cultures. We believe that this result represents a disconnect between individual and institutional trust, or a more fundamental disconnect between agent and human negotiation partners. It may be the case that Indians are more willing to trust virtual agents. Further extensions into varying the expressivity and levels of affect in our agents are currently underway in America, and future directions to include cross-cultural effects in Indian subjects would help to extend the ramifications of this result.

We found that H2 is correct, that Indians accept more offers, even to the extent of accepting poor offers below their BATNAs. Due to the nature of the ultimatum game and its high-risk outcomes, this result is expected for the predominantly prevention-focused Indian population. This led us to investigate trust, nationality, and acceptance rates. It was found that while nationality does strictly increase acceptance rates (per H2), this effect is partially mediated by trust levels.

We also found that Indians are insensitive to betrayal in their actions, but this betrayal is reflected in their later ratings of trust and fairness. Indians seem reluctant to display their disapproval by actively rejecting offers, but betrayals do affect measures of trust.

We found that H3 is true—Indians do find less integrative potential. However, due to the invalidation of H1, the direct correlation between trust and joint value found by Gunia seems to miss part of the picture. We believe it may be connected to the acceptance rates: Indians that "say YES" are not prepared to craft an offer of their own that leverages joint value, instead seeking the first harmonious outcome they find.

Finally, we found that H4 is true—all participants trust favor-seeking agents the most. Violating either through word or deed by failing to discuss favors or follow through on them reduces total levels of trust for the agent. While it is true that the main effect of H1 is shown here (Indians trust more in general), the pattern between Indians and US participants is similar—favor-seeking agents come out as the clear winners.

Thus, agents that utilize favor-language show a clear benefit for use in automated negotiations, even across cultures. Indian participants, traditionally thought of as having low individual trust, seem willing to trust virtual agents and automation, which is a compelling result for propagation and design of future agents. Additionally, these future agents must be able to unpack the distinction between acceptances that are made happily and those that have a negative result for measures of relationship trust. With these tenets in mind, the next generation of culturally and socially-aware agents can be built for effective negotiation.

REFERENCES

[1] Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C. M., Van den Bosch, K., & Meyer, J.-J. (2012). "Virtual reality negotiation training increases negotiation knowledge and skill". *12th International Conference on Intelligent Virtual Agents*.Santa Cruze, CA

[2] Core, M., Traum, D., Lane, H. C., Swartout, W., Gratch, J., Van Lent, M., & Marsella, S. (2006). "Teaching negotiation skills through practice and reflection with virtual humans". *Simulation, 82*(11), 685-701.

[3] Chalamish, M., & Kraus, S. (2012). "AutoMed: an automated mediator for multi-issue bilateral negotiations". Autonomous Agents and Multi-Agent Systems, 24(3), 536-564.

[4] de Melo, C. M., Carnevale, P., & Gratch, J. (2011, May). "The effect of expression of anger and happiness in computer agents on negotiations with humans" In The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3 (pp. 937-944). International Foundation for Autonomous Agents and Multiagent Systems.

[5] de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). "Reading people's minds from emotion expressions in interdependent decision making". Journal of personality and social psychology, 106(1), 73.

[6] DeVault, D., Mell, J., and Gratch, J. (2015). "Toward Natural Turn-Taking in a Virtual Human Negotiation Agent". In AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction. AAAI Press, Stanford, CA.

[7] Fulmer, C. A., Gelfand, M. J., Kruglanski, A. W., Kim-Prieto, C., Diener, E., Pierro, A., & Higgins, E. T. (2010). "On 'feeling right' in cultural contexts how person-culture match affects self-esteem and subjective well-being". *Psychological Science*.

[8] Gal, Y. A., Grosz, B. J., Kraus, S., Pfeffer, A., & Shieber, S. (2005, July). "Colored trails: a formalism for investigating decision-making in strategic environments". In *Proceedings of the 2005 IJCAI workshop on reasoning, representation, and learning in computer games* (pp. 25-30).

[9] Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007, January). "Creating rapport with virtual agents". In *Intelligent Virtual Agents* (pp. 125-138). Springer Berlin Heidelberg.

[10] Gunia, B. C., Brett, J. M., & Nandkeolyar, A. K. (2014). "Trust me, I'm a negotiator". Organizational Dynamics, 43, 27-36.

[11] Gunia, B. C., Brett, J. M., Nandkeolyar, A. K., & Kamdar, D. (2011). "Paying a price: culture, trust, and negotiation consequences". Journal of applied psychology, 96(4), 774.

[12] Halvorson, H. G., & Higgins, E. T. (2013). "Do you play to win--or to not lose?". *Harvard business review*, *91*(3), 117-20.

[13] Johnson, D., & Grayson, K. (2005). "Cognitive and affective trust in service relationships". *Journal of Business research*, *58*(4), 500-507.

[14] Kersten, G., & Noronha, S. (1999). "Negotiation via the World Wide Web: A cross-cultural study of decision making". Group Decision and Negotiation, 8(3), 251-279.

[15] Kraus, S., Hoz-Weiss, P., Wilkenfeld, J., Andersen, D. R., & Pate, A. (2008). "Resolving crises through automated bilateral negotiations". Artificial Intelligence, 172(1), 1-18.

[16] Lee, J. D., & See, K. A. (2004). "Trust in automation: Designing for appropriate reliance". Human Factors: The Journal of the Human Factors and Ergonomics Society, 46(1), 50-80.

[17] Madhavan, P., & Wiegmann, D. A. (2007). "Similarities and differences between human–human and human–automation trust: an integrative review". Theoretical Issues in Ergonomics Science, 8(4), 277-301.

[18] Mell, J., Lucas, G., & Gratch, J. (2015). "An Effective Conversation Tactic for Creating Value over Repeated Negotiations". In Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2015), Bordini, Elkind, Weiss, Yolum (eds.), May, 4–8, 2015, Istanbul, Turkey.

[19] Parasuraman, S., Singh, I. L., Molloy, R., & Parasuraman, R. (1992, September). "Automation-related complacency: A source of vulnerability in contemporary organizations". In *Proceedings of the IFIP 12th World Computer Congress on Education and Society-Information Processing'92-Volume 2-Volume 2* (pp. 426-432). North-Holland Publishing Co..

[20] Patton, B. (2005). Negotiation. The Handbook of Dispute Resolution, Jossey-Bass, San Francisco, 279-303.

[21] Peled, N., Gal, Y. A. K., & Kraus, S. (2011, May). "A study of computational and human strategies in revelation games". In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp. 345-352).

[22] Preacher, K. J., & Hayes, A. F. (2004). "SPSS and SAS procedures for estimating indirect effects in simple mediation models". Behavior research methods, instruments, & computers, 36(4), 717-731.

[23] Riegelsberger, J., Sasse, M. A., & McCarthy, J. D. (2005). "The mechanics of trust: A framework for research and design". International Journal of Human-Computer Studies, 62(3), 381-422.

[24] Rosenfeld, A., Zuckerman, I., Segal-Halevi, E., Drein, O., & Kraus, S. (2014, May). "NegoChat: a chat-based negotiation agent". In Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (pp. 525-532). International Foundation for Autonomous Agents and Multiagent Systems.

[25] Shrout, P. E., & Bolger, N. (2002). "Mediation in experimental and nonexperimental studies: new procedures and recommendations". Psychological methods, 7(4), 422.

[26] Van Kleef, G. A., De Dreu, C. K., & Manstead, A. S. (2004). "The interpersonal effects of emotions in negotiations: a motivated information processing approach". Journal of personality and social psychology, 87(4), 510.

[27] van Oostendorp, H, van der Spek, E.D. & Linssen, J.M. (2014). "Adapting the Complexity Level of a Serious Game to the Proficiency of Players". EAI Endorsed Transactions on Serious Games, 14(2),8.