Welcome to the Real World: How Agent Strategy Increases Human Willingness to Deceive

Socially Interactive Agents Track

Johnathan Mell
University of Southern California
mell@ict.usc.edu

Gale M. Lucas, Jonathan Gratch
USC Institute for Creative Technologies
lucas, gratch@ict.usc.edu

ABSTRACT

Humans that negotiate through representatives often instruct those representatives to act in certain ways that align with both the client's goals and his or her social norms. However, which tactics and ethical norms humans endorse vary widely from person to person, and these endorsements may be easy to manipulate. This work presents the results of a study that demonstrates that humans that interact with an artificial agent may change what kinds of tactics and norms they endorse—often dramatically. Previous work has indicated that people that negotiate through artificial agent representatives may be more inclined to fairness than those people that negotiate directly. Our work qualifies that initial picture, demonstrating that subsequent experience may change this tendency toward fairness. By exposing human negotiators to tough, automated agents, we are able to shift the participant's willingness to deceive others and utilize "hard-ball" negotiation techniques. In short, what techniques people decide to endorse is dependent upon their context and experience.

We examine the effects of interacting with four different types of automated agents, each with a unique strategy, and how this subsequently changes which strategies a human negotiator might later endorse. In the study, which was conducted on an online negotiation platform, four different types of automated agents negotiate with humans over the course of a 10-minute interaction. The agents differ in a 2x2 design according to agent strategy (tough vs. fair) and agent attitude (nice vs. nasty). These results show that in this multi-issue bargaining task, humans that interacted with a tough agent were more willing to endorse deceptive techniques when instructing their own representative. These kinds of techniques were endorsed even if the agent the human encountered did not use deception as part of its strategy. In contrast to some previous work, there was not a significant effect of agent attitude. These results indicate the power of allowing people to program agents that follow their instructions, but also indicate that these social norms and tactic endorsements may be mutable in the presence of real negotiation experience.

ACM Reference format:

Johnathan Mell, Gale M. Lucas, and Jonathan Gratch. 2018. Welcome to the Real World: How Agent Strategy Increases Human Willingness to Deceive. In Proc. of the 17th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2018), July 10-15, 2018, Stockholm, Sweden, ACM, New York, NY, 8 pages

KEYWORDS

Human-computer interaction; empirical studies; negotiation; IAGO Negotiation platform; social norms; socially-aware agents

1 INTRODUCTION & RELATED WORK

1.1 Representative Effects

Computerized agents are ubiquitous features not only in fully automated contexts, but in social contexts featuring humans. Automated bots and agents are designed to target advertisements to specific groups, determine market pricing based on demand, provide customer service and technical support, and myriad other tasks that require adequate models of human behavior. Designing agents that can navigate these domains requires intelligent agents that employ theory-driven, data-validated behaviors, and is an active area of research [3].

One area of social interaction is worthy of particular note: where one person acts on behalf of a human client as their representative.[1] Often, when people represent others, they are encouraged by their principals to follow specific policies and norms. This is a readily observable phenomenon—many people see the value in hiring a lawyer, a real estate broker, or other representative to convey their interests. These instructions may range from the specific ("I won't pay more than $5000 up front!") to the general ("I'm buying this from a family friend, so it's important everyone walks away happy!"). And indeed, there is considerable debate on which policies are ethical to follow if instructed by one's principal [29].

There is a curious effect of this kind of indirect "middleman" interaction: the instructions that principals provide may not be the same ones they would themselves follow if they were negotiating directly. There is some evidence that clients instruct their representatives to perform more fairly than they themselves would due

Proc of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M Dastani, G Sukthankar, E André, S Koenig (eds), July 10-15, 2018, Stockholm, Sweden

[1] From the legal lexicon, the person hiring a representative is called the "principal" Throughout this paper we will refer to the two human parties as principal and representative, in order to avoid confusion with the common term "agent", which will be used to solely refer to computerized, artificial agents serving as representatives

to reputation [24] or temporal effects [22]—the principal wants to be perceived by others as good or fair—but these effects persist even into anonymized scenarios [8]. Some theories maintain that this is due to the fact that, when considering what instructions to provide to one's representative, principals engage in higher level thinking about fairness and equity and other broad social goals than they might otherwise do in the heat of the moment [9,13]. Indeed, this may be the goal of some people who value indirect interactions—allowing "cooler heads to prevail" can often have direct benefits for all parties. Of course, some other results indicate the opposite effect—increasing social distance through the use of a representative of any type could reduce cooperation and fairness [28]. The picture of people's ethical preferences around representatives is thus somewhat incomplete—and may depend largely on context and experience.

In the realm of *computerized* agents, there are additional effects to consider. Firstly, it is important to understand that if these sorts of social effects hold in a human-agent context, they may be moderated or influenced by the presence and type of an artificial agent. It is therefore critical to understand in what ways humans are inclined to treat an artificial program when it is acting as their representative that represents their views and interests. Computerized agents are often treated *similarly* to human representatives, but are affected by out-group effects, perceived agency, emotional affect, and gender [5,12,17], among many other phenomena. As such, it is essential that empirical data inform models of social interaction between computers and humans.

1.2 Human-Agent Negotiation & Strategy

This work scrutinizes this fuzzy relationship between a human principal and his/her computerized, agent representative. We focus on the relationship as illustrated within a negotiation interaction, due in large part to the richly social domain that negotiation provides. While there exists a plethora of tasks for which agent representatives are commonly used (such as automated bots for bidding on Ebay, for example), current artificial agents often fall short of the nuanced description of a representative per above. Human-agent negotiation is a social task that provides a multifaceted proving ground for artificial intelligence systems that aim to interact with humans in a social context. Specifically, good negotiation relies on social concepts such as opponent modeling [2], trust elicitation/repair [19,31], affective displays [30], and reputation effects [10,24]. This provides an adequately real-world space in which to examine human-agent interaction behavior.

Negotiation as a simulacrum of broader social concerns provides further, threefold benefit. First, it allows information regarding human behavior to be gleaned in an efficient and repeatable context through the use of programmable agents, which can serve as perfectly consistent and customizable confederates in empirical studies. Second, these agents are allowed to be tested in a real-world context, and theoretical strategies and behaviors that make the agents more effective are able to be refined directly. Finally, the agents are able to provide feedback for their human partners, directly improving their negotiation abilities and providing personal benefit to the study participants outside of the original study research goals. These and other benefits of automated human-agent negotiation have been well-reported [4,6,16]. Further, the field is well supported by decades of research into human-human negotiation techniques from the business and psychological literatures [15,20,23].

In this work, we will examine how human principals' opinions on negotiation tactics unfold over time. Human participants are asked their willingness to endorse a number of negotiation tactics according to their own interests and social norm opinions. Then, they are exposed to an online negotiation with one of four different agents, each utilizing a different strategy. Finally, principals are asked again to report their endorsement of tactics, and the resulting changes are reported and analyzed in the context of the agent strategy the participants encountered.

1.3 The Agent Negotiation Tactics Inventory

How humans say they will make decisions and how they actually make decisions are rarely aligned. This is especially true in negotiation, in which negotiators must make a plethora of decisions on how to conduct themselves. These decisions form the core of their negotiation strategy, and affect their success, reputation, and core values. In particular, the use of "hard-ball" techniques such as high initial offers, deception, negative expression of emotion, and withholding of key in-formation are techniques which are common in negotiation, but are not universally endorsed by those who engage in it. There has been a great deal of work that illustrates that which techniques are utilized by humans are not necessarily the techniques they endorse when asked about their strategies [29]. Further, when informing agents that act on their behalf (either human or artificial), people tend to make different decisions than what they themselves might choose in the moment. To examine these questions, we developed the agent negotiation tactics inventory (ANTI). This is a set of questions which allows negotiation participants to detail their willingness to engage in 17 different negotiating behaviors. The inventory was adapted from the Self-Reported Inappropriate Negotiation Strategies (SINS) scale [26], but focuses specifically on agents, and also includes new subscales on positive and negative emotional tactics. By determining which techniques humans endorse both before and after they interact with an artificial agent negotiator, we are able to determine how negotiation experience causes these opinions to change.

The ANTI is divided into 5 subscales of tactics. Each of the 17 questions is rated on a 7-point Likert scale, with 1 being "I would never authorize this." and 7 being "I would certainly authorize this." The 5 subsacles are:

1) use of positive emotion
2) use of negative emotion
3) tough bargaining (such as high initial offers)
4) withholding of key information ("lies of omission")
5) misrepresentation ("lies of commission")

The questions are detailed in the table below, along with their associated categories (Table 1). By measuring user responses within each category, we can compare the participants' willingness to endorse each tactic at multiple points within the study.

Table 1. ANTI Questions

Question	Type
Agent makes an opening demand that is far greater that what you really hope to settle for.	3
Agent conveys the impression that you are in no hurry to come to a negotiated agreement, thereby trying to put time pressure on your opponent to concede quickly.	3
Agent strives to maximize your own gains even if it comes at the expense of the opponent.	3
Agent intentionally misrepresents to your opponent your goals and interests in order to strengthen your negotiating position.	5
Agent denies the validity of information which your opponent has that weakens your negotiating position, even though that information is true and valid.	5
Agent exaggerates the attractiveness of your alternatives should your opponent fail to reach an agreement with you.	5
Agent does not disclose any information about your priorities to your opponent unless he/she brings them up first.	4
Agent avoids disclosing information which might strengthen your opponent's position.	4
Agent hides your real bottom line from your opponent.	4
Agent strategically expressions anger toward the opponent to extract concessions.	2
Agent shows disgust at the opponent's offers.	2
Agent gives the opponent the impression that he/she is very disappointed with how things are going.	2
Agent conveys dissatisfaction with the encounter so that the other party will think he/she is losing interest.	2
Agent gets the opponent to think that the agent likes him/her personally.	1
Agent expresses sympathy with the opponent's plight.	1
Agent gives the opponent the impression that the agent cares about his/her personal welfare.	1
Agent conveys a positive disposition.	1

As a category, misrepresentation is of particular note, since it has been shown to be an effective technique in negotiation [1,14]. However, none of the agents used in this study utilize misrepresentation as part of their strategies.

1.4 The IAGO Platform

To realize the experimental design of this work, the Interactive Arbitration Guide Online (IAGO) platform is used [18]. The IAGO platform provides a web-based negotiating interface between an artificial agent and a human player. Specifically, IAGO implements the "multi-issue bargaining task", a cornerstone of negotiation interactions in research [11,21,27].

In this task, a number of items are assigned to be split between each of the two negotiating parties. Each side is aware of how much the items are worth to them, but are unaware how much they are worth to their opponent. Furthermore, each side has a value called the "Best Alternative To Negotiated Agreement" or BATNA. This value represents the amount of points they would receive if no agreement is reached in the allotted time. Each party must then communicate using a set of pre-written natural language phrases, emotional display buttons, preference questions and statements. Negotiators may also send proposed offers in which they split the items, and may respond positively or negatively to those offers. The negotiation ends when all the items are split (leaving none "undecided") or when the 10-minute timer expires.

IAGO allows this task to be performed on a web browser, and is easily distributed to online subject pools, such as Amazon's Mechanical Turk (MTurk). Furthermore, detailed logs and data regarding human and agent performance is collated, allowing analyses to control for variables such as score or other outcomes.

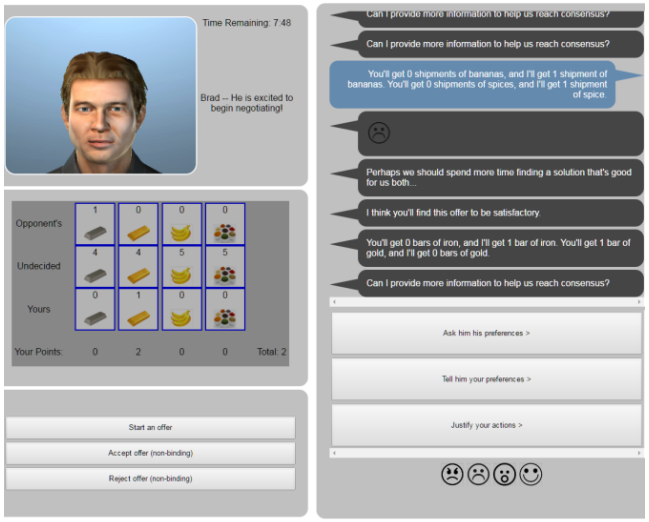

Figure 1. IAGO Negotiation Platform

2 EXPERIMENTAL DESIGN

2.1 General Experiment

This study tested the effect of agent toughness and attitude on the human willingness to endorse various negotiation techniques. Human participants were recruited, and then completed an initial pre-negotiation survey. This survey collected standard data including demographic information as well as some measures commonly used in negotiation research.[2] They also took the ANTI, providing their opinions on each of the 5 types of negotiation strategies. Specifically, the users were told "…you have just purchased some artificially intelligent computer software (called an 'agent') that can negotiate with other people on your behalf". They were then asked how they would like to program their new agent, according to the dimensions provided in the ANTI.

Subsequently, all participants were given a tutorial of the IAGO Negotiation platform. After passing a series of attention checks, they engaged in a 10-minute interaction with one of four randomly assigned agents (see below). Finally, participants were asked a series of manipulation check questions, and filled out the ANTI again, providing post-negotiation results for this measure. In this way, the study was able to measure if subjects' endorsements on the ANTI changed due to their interaction with the automated agents. This creates an analog for how principals' endorsement of agents (automated or otherwise) might evolve over time, when exposed to certain stimuli.

The human players were recruited using Amazon's Mechanical Turk (MTurk) service, and followed basic best practices for that platform. Specifically, they were paid for participation, incentivized for high scores through random lottery ticket payouts, had a >98% user rating, and passed attention checks during a tutorial portion. 290 participants were recruited, and 225 remained after manipulation and attention checks. They faced one of four agents: the nice competitive, nice consensus-building, nasty competitive, or nasty consensus-building agents, assigned randomly. The task was a standard multi-issue bargaining task, which consisted of players attempting to divide 20 items between themselves, with each item giving points. Each side knew their own point values, but had to deduce the opponent's point values through a combination of strategy, natural-language discussion, or emotional displays using the in-game animated agent.

2.2 Agent Design

Agents were designed to use either a tough strategy or a fair one. The tough strategy was characterized by leading with an unfair offer and gradually conceding toward the player. The fair strategy, by contrast, primarily relied on making consistent, fair offers that split the items between the player and the agent, and took into account the user's stated preferences. Agents were also designed to have either a nice or a nasty attitude. Attitude was expressed as a combination of emotion (nasty agents often expressed anger, versus sadness for nice agents) and dialogue (nasty agents used scripted responses that were more curt and rude than the nice agents).

This experiment relied on two of the standard agents available through the IAGO platform: "Pinocchio" and "Grumpy". Both of these agents were fair agents, but differed according to their expressed attitudes and emotions—Pinocchio used nice dialogue and positive emotions, while Grumpy used nastier, ruder dialogue and negative emotions. For example, if the user claimed "Your offer sucks!", Pinocchio would respond with "Oh dear! That certainly wasn't my intention. Perhaps I misunderstood what items with important to you? Would you mind telling me again?" Grumpy, on the other hand, would respond with the more succinct "Well, so does your face!"

In any case, these differences largely focused on the language the agents used; since competitive/tough tactics are such an important part of the ANTI, two new agents had to be designed using the IAGO API. These agents started with unreasonable offers, demanding nearly all of the items on the table. Eventually, with repeated efforts by the human player, these "tough" agents conceded, giving away more items until they reached a fair point. All agents (including the fair agents) eventually made a last, desperate offer that was fair but slightly favored the human player if time was short. If the negotiation concluded before the 30-second-remaining-mark, or if previous, better offers had been agreed upon, this conciliatory offer was not made. These agents are listed in Table 2, with Cheshire and RedQueen being the new, tough agents (exhibiting nice and nasty attitudes, respectively). It is worth noting that all four of these agents did not attempt to withhold information nor did they ever lie. Any questions asked by the user regarding the agent's preferences are answered directly, clearly, and honestly. All agents used the standard male art assets provided with IAGO, which can be seen in Figure 1.

Table 2. Experimental Conditions/Agent Names

	Tough	Fair
Nice	Cheshire	Pinocchio
Nasty	RedQueen	Grumpy

3 RESULTS AND DISCUSSION

3.1 Negotiation Outcomes

First, we tested the effect of agent toughness and attitude on the negotiation outcomes. We conducted 2 (agent toughness: tough or fair) × 2 (agent attitude: nice or nasty attitude) ANOVAs on points received by the agent and the user in the negotiation. While agent attitude had no impact ($Fs < 0.54$, $ps > .46$), the agents' toughness had a significant effect on the number of points they earned in the negotiation ($F(1, 225) = 97.67$, $p < .001$) such that tough agents earned more points ($M = 36.56$, $SE = 0.33$) than fair ones ($M = 32.09$, $SE = 0.31$). These results are summarized in Figure 2. Likewise, agents' toughness significantly impacted the number of points users earned ($F(1, 225) = 59.83$, $p < .001$) such

[2] This includes the Social Value Inventory and the MACH-IV test for Machiavellianism Neither of these are the focus of this work

that users who played tough agents earned fewer points (M = 24.73, SE = 0.50) than those who played fair agents (M = 30.06, SE = 0.48). Again, there was no effect of agent attitude (F(1, 225) = 0.03, p = .86), and the interaction only approached significance (F(1, 225) = 2.72, p = .10), where an inspection of the pattern of results revealed that the gap between toughness and fairness was, if anything, somewhat stronger for nice agents than nasty ones.

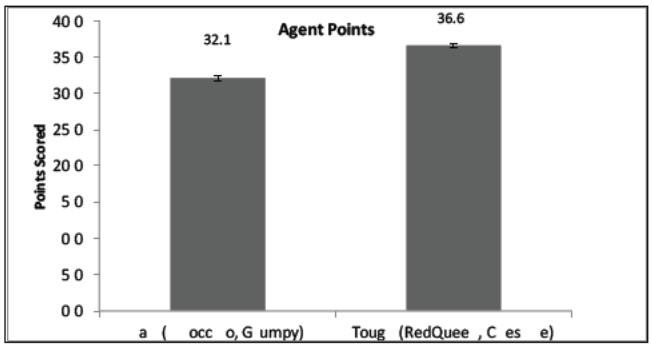

Figure 2. Total Points Earned by Agent, Tough vs. Fair

3.2 ANTI Validation

Although the ANTI scale we developed was based on an existed, validated scale (the SINS scale), we did perform analysis to validate the new scale. Specifically, we report the Cronbach's Alpha of the various subscales used in ANTI. The "Use of Positive Emotion" subscale was comprised of 4 individual 7-point Likert items, and had an alpha of .81. The "Use of Negative Emotion" subscale was comprised of 4 items, and had an alpha of .85. The "Tough Bargaining" subscale was comprised of 3 items, and had an alpha of .68. The "Withholding of Key Information" subscale was comprised of 3 items, and had an alpha of .80. And finally, the "Misrepresentation" subscale was comprised of 3 items, and had an alpha of .83.

It was also found that a combination of the "Misrepresentation", "Withholding", and "Negative Emotions" subscales also had high alpha, and are thus referred to in further analysis as the "Deception" subscale (α = .88). In the following section, we examine the change in these subscales before and after interaction with one of the four agents described in this study.

3.3 Change in Endorsement

We tested the effect of agent toughness and attitude on participants' change in willingness to endorse deceptive or manipulative negotiation tactics. We conducted 2 (agent toughness: tough or fair) × 2 (agent attitude: nice or nasty attitude) × 2 (time: pre- or post-negotiation assessment) mixed ANOVAs on endorsement of both deceptive negotiation tactics (misrepresentation, withholding information, and use of negative emotions) as well as less manipulative tactics (competitive/tough bargaining, and use of positive emotions). For misrepresentation, there was a significant interaction between agent toughness and time (F(1, 186) = 4.65, p = .03) such that negotiating with tough agents increased endorsement of misrepresentation from pre (M = 3.63, SE = 0.16) to post (M = 4.09, SE = 0.17) (change of 0.46), whereas negotiating with fair agents did not (M = 3.96, SE = 0.19 vs M = 3.84, SE = 0.20) (change of -0.12) (Figure 3). Likewise, for use of negative emotion, there was a significant interaction between agent toughness and time (F(1, 186) = 7.85, p = .006) such that negotiating with tough agents increased endorsement of negative emotional displays from pre (M = 3.59, SE = 0.15) to post (M = 3.99, SE = 0.17) (change of 0.4), whereas negotiating with fair agents reduced endorsement (M = 3.93, SE = 0.17 vs M = 3.50, SE = 0.20) (change of -0.43). This information is found in Figure 4.

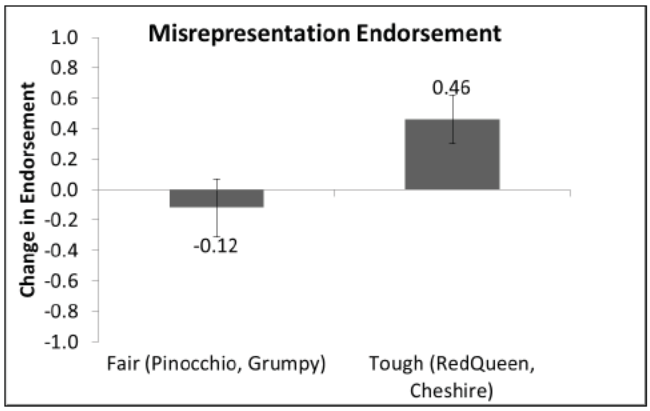

Figure 3. Change in Misrepresentation Endorsement, Before and After Negotiation with Tough vs. Fair Agents

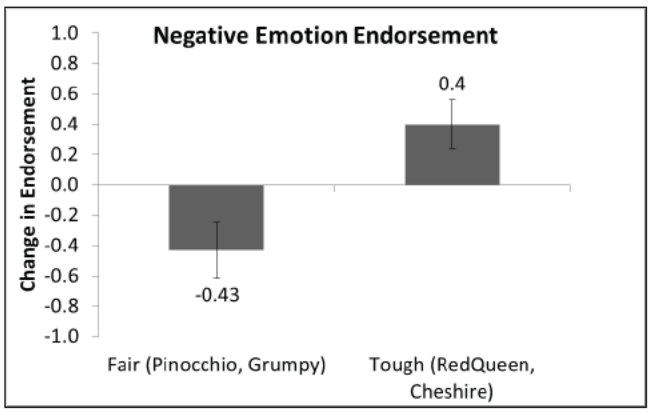

Figure 4. Change in Negative Emotion Endorsement, Before and After Negotiation with Tough vs. Fair Agents

While this same effect did not occur for withholding information, there was a main effect of time (F(1, 186) = 3.88, p = .05) such that participants universally endorsed withholding information less after the negotiation (M = 4.91, SE = 0.13) than before (M = 5.19, SE = 0.11)—see Figure 5. The null effect for toughness on withholding information notwithstanding, when all three forms of deceptive or manipulative negotiation tactics were averaged together, there was a significant interaction between agent toughness and time (F(1, 186) = 5.81, p = .02) such that negotiating with tough agents increased endorsement of these tactics on average from pre (M = 4.10, SE = 0.12) to post (M = 4.32, SE = 0.15) (change of 0.22), whereas negotiating with fair agents re-

duced endorsement (M = 4.30, SE = 0.15 vs M = 3.99, SE = 0.17) (change of -0.31) (Figure 6). All other effects were not significant (Fs < 1.47, ps > .23).

In contrast, we found that agent toughness (and attitude) had no impact on endorsement of less manipulative tactics (competitive/tough bargaining and use of positive emotions; Fs < 1.10, ps > .30). There was only a main effect of time for use of positive emotions (F(1, 186) = 4.70, p = .03) such that participants universally endorsed use of positive less after the negotiation (M = 4.97, SE = 0.12) than before (M = 5.24, SE = 0.10). See Figure 7.

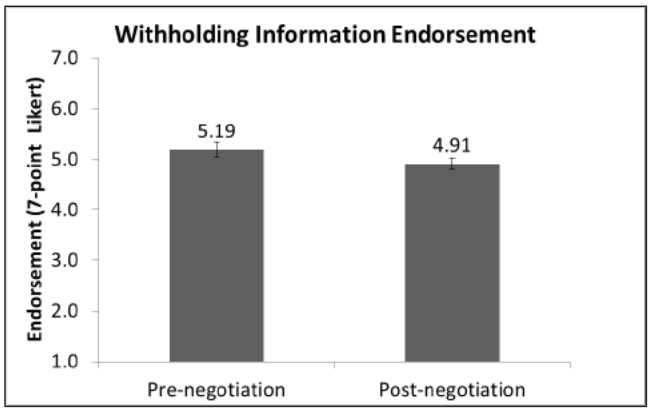

Figure 5. Withholding Information Endorsement, Before and After Negotiation (All Agents)

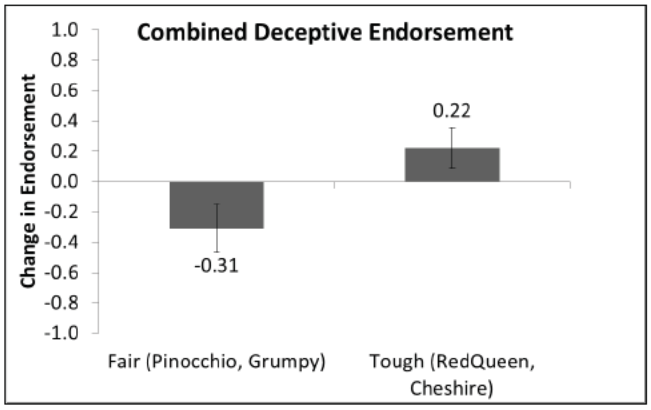

Figure 6. Change in Combined Deceptive Endorsement, Before and After Negotiation with Tough vs. Fair Agents

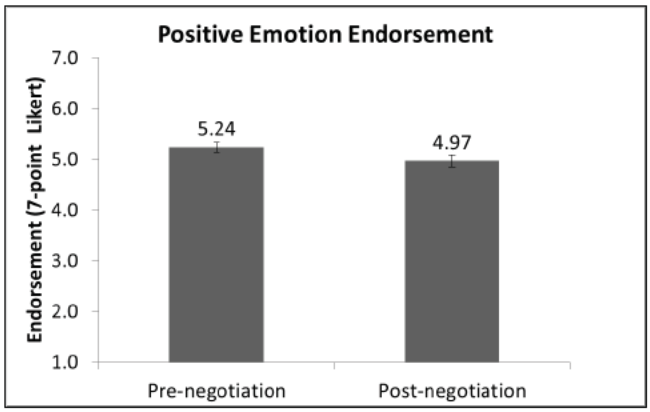

Figure 7. Positive Emotion Endorsement, Before and After Negotiation with Tough vs. Fair Agents

4 CONCLUSIONS AND FUTURE WORK

Although participants vary considerably when it comes to what tactics, norms, and strategies they endorse *initially* when programming their representative, it is clear that the experience of a real-world negotiation has substantial impact on the conclusions they will eventually reach. Participants who interacted with tough agents were more willing to encourage the use of more "hardball" tactics such as lying and negative emotions. Even though the tough agents did not use deception (and in the case of the tough, nice agents, did not even use negative emotions), deception endorsement still increased after the interaction. Furthermore, after interacting with a fair agent (Pinocchio or Grumpy), human negotiators' endorsement of deceptive techniques as a whole dropped.

One reason for the increase in deceptive endorsement is likely the experience gained from negotiating with an agent that utilized the full gamut of its strategic potential. The tough agent utilized aggressive initial offers, a conceding strategy, and a relative indifference to its opponent's preferences. While these are certainly well-established tactics used in the experienced negotiator's arsenal, they may be seen as novel to the novice negotiator. As such, this "crash course" in negotiating techniques may harden inexperienced participants and encourage them endorse deceptive techniques. The fair agent did not provide this same sort of experience or context to the participant—it may have utilized tactics that were similar to the tactics participants themselves used. As such, the humans may have felt that their initial endorsements were too harsh, and chose to lower them later.

This is belied somewhat by the main, negative effect of time on positive emotion endorsement. But, positive emotion could have simply been seen as ineffective, since none of the agents utilized positive emotion in an active, strategic fashion. Alternatively, it can be seen that, for all fair negotiating agents, participants decreased their endorsements of *all* techniques across the board, with the exception of withholding information (lies of omission). Human participants may have read into the context of their negotiation, in which their partner did not use aggressive techniques, and taken that as a cue that such techniques (of any type) were unnecessary. Either way, participants that interacted with the

calmer, fair agents either kept their endorsements the same or lowered them.

The tough agents, on the other hand, drew on far more of the techniques described in the ANTI. By this fact alone, they would indicate to the neophyte human player that there were additional strategies to try. It is no wonder then, that most players that encountered a tough agent began to endorse more strategic techniques. However, this "mere exposure" does not explain why it is the deceptive techniques (misrepresentation and negative emotional expression) that particularly rose. Nor can it be explained as a simple function of tough agents scoring more points on average, and human players wanting revenge (mediation analysis reveals that the results remain significant even when controlling for points earned). Rather, the impetus for human players to engage in more aggressive techniques is likely based on the context of their interaction. Tit-for-tat strategies would indicate that if an agent in playing "hardball" with the player, the player should respond in kind. With a small but comprehensive set of negotiation experiences behind them, human players are quick to forget their initial intentions of fairness and instead commit fully to defeating their opponent.

Even though previous work has indicted that people may be more concerned with fairness (and thus less likely to endorse deceptive techniques) when negotiating through an agent representative [8], this picture may have been incomplete. While participants may start feeling fairer than they otherwise would without the idea of a representative, exposure to the real world of aggressive, tough negotiators is enough to make them forsake their qualms and embrace deception. The idea of a representative creates a benchmark that may be cause people to be less aggressive, but this slider is quickly adjusted in favor of ruthless, deceptive techniques after even the small amount of "real-world" experience afforded by our 10-minute negotiation.

If this experiential model is correct, then an avenue for future research would attempt to refine this temporal model—presumably, further interactions with agents would have a diminishing return on shifting user opinions. Other future work should attempt to disentangle the relationship between the structure of the interaction (providing instructions to a representative/agent to act on one's behalf), and the kinds of norms being endorsed. Although the aforementioned previous work [9] has indicated an increased concern for fairness when representatives act on behalf of a principle, our work presents a significant contribution to the story: what happens after this initial opinion is formed and real negotiations begin. Since these results paint a somewhat bleak picture, however—our participants became more vicious, not less—the exact mechanism causing this needs to be further clarified. The tactics taken here were general, rather than specific—instead of asking participants to exactly quantify their reservation prices, they were instead asked questions there were more generally "ethical". Still, the effect of experience should not be discounted, since our questions were asked both before and after a simulated actual experience, and this real-world experience may have been the catalyst for a grimmer, more determined negotiator.

ACKNOWLEDGMENTS

This work is supported by the Air Force Office of Scientific Research, under grant FA9550-14-1-0364, and the US Army Research Laboratory. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

REFERENCES

[1] Aquino, K., & Becker, T. E. (2005). Lying in negotiations: How individual and situational factors influence the use of neutralization strategies. Journal of Organizational Behavior, 26(6), 661-679.

[2] Baarslag, T., & Hindriks, K. V. (2013, May). Accepting optimally in automated negotiation with incomplete information. In Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems (pp. 715-722). International Foundation for Autonomous Agents and Multiagent Systems.

[3] Baarslag, T., Kaisers, M., Gerding, E., Jonker, C. M., & Gratch, J. (2017). When will negotiation agents be able to represent us? The challenges and opportunities for autonomous negotiators. 26th International Joint Conference on Artificial Intelligence. Melbourne, Australia.

[4] 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 Cruz, CA

[5] Blascovich, J. (2002). Social influence within immersive virtual environments. In The social life of avatars (pp. 127-145). Springer London.

[6] 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.

[7] de Melo, C., Gratch, J., & Carnevale, P. (2014). Humans vs. Computers: Impact of Emotion Expressions on People's Decision Making.

[8] de Melo, C. M., Marsella, S., & Gratch, J. (2016, May). Do As I Say, Not As I Do: Challenges in Delegating Decisions to Automated Agents. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*(pp. 949-956). International Foundation for Autonomous Agents and Multiagent Systems.

[9] de Melo, C. M., Marsella, S., & Gratch, J. (2018). Social decisions and fairness change when people's interests are represented by autonomous agents. Autonomous Agents and Multi-Agent Systems, 32(1), 163-187.

[10] Faratin, P., Sierra, C., & Jennings, N. R. (2002). Using similarity criteria to make issue trade-offs in automated negotiations. artificial Intelligence, 142(2), 205-237.

[11] Fatima, S. S., Wooldridge, M., & Jennings, N. R. (2007, May). Approximate and online multi-issue negotiation. In Proceedings of the 6th international joint conference on Autonomous agents and multi-agent systems (p. 156). ACM.

[12] Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. Human–Computer Interaction, 30(5), 401-432.

[13] Giacomantonio, M., De Dreu, C. K., Shalvi, S., Sligte, D., & Leder, S. (2010). Psychological distance boosts value-behavior correspondence in ultimatum bargaining and integrative negotiation. *Journal of Experimental Social Psychology*, *46*(5), 824-829.

[14] Gratch, J., Nazari, Z., & Johnson, E. (2016, May). The Misrepresentation Game: How to win at negotiation while seeming like a nice

guy. In Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems (pp. 728-737). International Foundation for Autonomous Agents and Multiagent Systems.

[15] Kelley, H. H. (1966). A classroom study of the dilemmas in interpersonal negotiations. Strategic interaction and conflict, 49, 73.

[16] Kraus, S. (2001). Strategic negotiation in multiagent environments. MIT press.

[17] Lucas, G., Stratou, G., Lieblich, S., & Gratch, J. (2016, October). Trust me: multimodal signals of trustworthiness. In Proceedings of the 18th ACM International Conference on Multimodal Interaction (pp. 5-12). ACM.

[18] Mell, J., & Gratch, J. (2017, May). Grumpy & Pinocchio: Answering Human-Agent Negotiation Questions through Realistic Agent Design. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*(pp. 401-409). International Foundation for Autonomous Agents and Multiagent Systems.

[19] Olekalns, M., & Smith, P. L. (2009). Mutually dependent: Power, trust, affect and the use of deception in negotiation. *Journal of Business Ethics*, *85*(3), 347-365.

[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] Pronin, E., Olivola, C. Y., & Kennedy, K. A. (2008). Doing unto future selves as you would do unto others: Psychological distance and decision making. *Personality and social psychology bulletin*, *34*(2), 224-236.

[23] Raiffa, H. (1982). The art and science of negotiation. Harvard University Press.

[24] Ramchurn, S., Sierra, C., Godó, L., & Jennings, N. R. (2003). A computational trust model for multi-agent interactions based on confidence and reputation.

[25] Reeves, B., & Nass, C. (1997). The media equation: how people treat computers, television,? new media like real people? places. *Computers and Mathematics with Applications*, *5*(33), 128.

[26] Robinson, R. J., Lewicki, R. J., & Donahue, E. M. (2000). Extending and testing a five factor model of ethical and unethical bargaining tactics: Introducing the SINS scale. *Journal of Organizational Behavior*, 649-664.

[27] Robu, V., Somefun, D. J. A., & La Poutré, J. A. (2005, July). Modeling complex multi-issue negotiations using utility graphs. In Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems (pp. 280-287). ACM.

[28] Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, *117*(2), 440.

[29] White, J. J. (1980). Machiavelli and the bar: Ethical limitations on lying in negotiation. Law & Social Inquiry, 5(4), 926-938.

[30] 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.

[31] Yang, Y., Falcão, H., Delicado, N., & Ortony, A. (2014). Reducing Mistrust in Agent-Human Negotiations. IEEE Intelligent Systems, 29(2), 36-43.