Results of the First Annual Human-Agent League of the Automated Negotiating Agents Competition

Full Paper

<Authors Redacted for Review>

ABSTRACT

In designing agents that interact with humans, human-agent negotiation has emerged as a challenge problem for furthering research into a number of related fields. Specifically, human-aware agents are capable of teaching and training humans, and can provide valuable insight into human cognition. To further the state-of-the-art of these types of agents, a competition was held in conjunction with the 8th annual Automated Negotiating Agent Competition (ANAC). We present the results of the first annual Human-Agent League of ANAC—a departure from previous ANACs, which have historically focused on similar types of interactions, but in the agent-agent context.

By introducing a new human-agent negotiating platform to the researching community at large through association with ANAC, we facilitated new advancements in human-aware agents. This has succeeded pushing the envelope in agent design, and creating a corpus of useful human-agent interaction data [1]. Our results indicate a variety of agents were submitted to the competition, and that their varying strategies had distinct outcomes on many measures of the negotiation. These agents approach the problems endemic to human negotiation, including user modeling, bidding strategy, rapport-building techniques, and strategic bargaining. Some agents which employed advanced tactics in information gathering or emotional displays were able to gain more points than their opponents, while others were considered more "likeable" by their partners, and were given more positive reviews.

We examine these results, and provide an overview herein.

KEYWORDS

Human-Agent Negotiation; IAGO Negotiation Platform; Empirical Studies; Human-Computer Interaction

1 INTRODUCTION & BACKGROUND

1.1 Negotiation as a Challenge Problem

Automated negotiation has been presented in previous works as a key challenge problem for the advancement of virtual humans—agents that possess human-like characteristics. Negotiation has been described as an "indispensable skill for any social creature", and the fruits of research into automated negotiating agents have myriad benefits [7]. Specifically, automated negotiating agents are capable of training and teaching humans to be better negotiators, and insights gleaned from the design of such agents can lead to the development of more general teaching and training agents [4,10]. Additionally, automated agents provide benefits to empirical studies, as they can provide a consistent confederate, unchanging and tireless over the course of a human-subjects study [7].

Beyond training humans to be better negotiators, negotiating agents can also serve as virtual assistants for a variety of applications. Google's recent Duplex demo demonstrates this astutely, as an agent acts on behalf of a human to create an appointment, negotiating with a human partner to decide on an agreeable time [9]. Legal scholarship has long examined the ethics of acting on behalf of others [13,19], and psychology and computer science have both explored the mechanisms by which humans instruct their representatives, be they human or virtual [5, 21].

Of course, such negotiating agents require a number of components to render them fully capable of acting as negotiating partners. The design of negotiating agents presents problems in strategy, opponent modelling, preference elicitation, rapport-building, natural language generation/understanding, non-verbal behavior generation, use of emotional affect, to name just a few. When creating agents, designers must address, at the bare minimum, how an agent will model its opponent (through their utterances and/or offer patterns), how the agent will make offers, and how it will respond to offers. These problems are well-summarized in the literature, and become even more critical when considering negotiating agents not just in their role as partners, but as representatives or proxies for actual human negotiators [7].

1.2 Negotiating Agents in Competition

Designing automated negotiating agents has been an ongoing area of research. One such way in which this research has been driven is by researcher collaboration during the Automated Negotiating Agents Competition (ANAC) [8]. ANAC, as well as individual research efforts, have focused on designing agents that can interact with each other in a variety of economically-styled games, including the multi-issue bargaining task (a standard task). These agents attempt to model opponent preferences, adapt bargaining strategies, and provide optimal offers over a variety of constraints.

While ANAC has been a recurrent, successful competition for 8 years (2018 will mark the 9th annual ANAC), it has been focused primarily on agent-agent negotiation. Human-agent negotiation is fundamentally different than agent-agent negotiation, and the Human-Agent Track of ANAC was added in

[1] This data will be released in conjunction with this paper Further information is available from the first author, <redacted for anonymous submission>

order promote further research into this promising area. Agents developed with human factors in mind can provide training and education to human users. This is of particular use in negotiation, since training negotiation is often a lengthy and expensive process, often requiring years of study and/or professional training programs offered through business schools or the like. Furthermore, studies backed by human data allow for agents to be developed that are more emotive, more realistic, and more helpful in myriad human-agent contexts outside of negotiation.

Agents that are developed with humans in mind need not emulate their behaviors (although that may be a goal in certain circumstances), but do need to be able to respond human behaviors intelligently. First and foremost, this presents a problem for agents to accurately model their opponents. Since negotiation not a fully-visible scenario, agents must make educated decisions about user preferences, user alternatives to agreement (referred to as "BATNA"), and individual user personality types/strategies. Agents may use a variety of information sources to come to their conclusions, including the user history of offers, the natural language utterances they send, and even the emotions they express.

Beyond opponent modeling, agents must decide their own behaviors in a negotiation. Although starting with a tough offer and conceding slowly a tried and true negotiating tactic in both human and agent-agent negotiations, it does have its risks. Certain negotiating partners are more disagreeable than others, and may fight back, lowering the joint value possible if an agent is too aggressive early on. Agents may also decide how optimistic or pessimistic to be in the face of uncertainty. And they must decide how to utilize their BATNA strategically, and if they want to follow certain strategies that may be considered unethical (such as lying).

Finally, agents must also be able to adapt to individual differences in negotiation. While a single strategy may be helpful in many cases, the best agents (and negotiators of any provenance) must be able to read the situation and adjust accordingly. These problems for virtual agents may be addressed with a variety of solutions, many of which are on display here, in the results from ANAC 2017.

1.3 Agents & Negotiation Strategies

As with human-human negotiation, there are many effective tactics that can lead to success in agent negotiation. Among the agents that were submitted to this competition, for example, there are several that use emotion in an attempt to influence their opponent. This strategy (particularly the use of negative emotions to gain concession) has been well-documented both in human-human and human-agent contexts [20]. Agent Cena, the runner-up agent in terms of points scored, applied this tactic.

Conversely, some agents in the competition attempted to build rapport with the human-participant through the use of positive emotion, with the hope that it would lead to greater value [rapport ref]. The top-rated agent for likeability, Agent Wotan, used this strategy.

Another major tactic used in the competition was deception. The aptly-named LyingAgent was able to employ this strategy to "grow the pie" of the negotiation and claim more than its fair share of the result. Lying in negotiation is a well-established tactic, and indeed the LyingAgent has been subsequently described in detail in publications following the competition [2, 17,18]. The ethics of these tactics have also been explored in some detail [19].

All the agents also model their opponent to an extent, although Agent Murphy and Elphaba take the greatest steps in doing so, both incorporating some kind of reliability metric for how certain they are of opponent preferences. Agent Murphy, additionally, takes a pessimistic view of the potential outcomes.

1.4 Submitted Agents

1.4.1 Pinocchio (Baseline)

The baseline agent was provided as source code to all participants of the competition. Pinocchio followed a straightforward strategy—it attempted to gain information about the human's preferences, then made fair partial offers, giving away one item at a time until all items were assigned. These offers were made in response to user offers, and Pinocchio did not generate offers on its own outside of these counter-offers. It also used only positive emotions. Pinocchio has been discussed in detail in previous work, as it is a part of the IAGO standard toolkit [11].

1.4.2 Elphaba

Elphaba used a mixture of positive and negative emotions (anger, happiness, and sadness), depending on a user "reliability" score. Reliability increased when user preferences were detected, and decreased when contradictions in user statements were exposed. From this, Elphaba attempted to grow value for both the human and the agent.

1.4.3 Agent Cena

Agent Cena was an aggressive agent, utilizing tough statements and negative emotions to attempt to browbeat opponents into giving up value. Cena started with unreasonably high offers (90% of the total value), and slowly conceded, accelerating in the last 3 minutes of the negotiation. Cena did attempt to give the human player their most valuable item first, however.

1.4.4 Agent Wotan

Wotan was similar to Cena in that it started with a high offer and conceded, although it would concede further and faster than Cena. Furthermore, Wotan only made full offers (in which all items were assigned to either the player or the agent), in an attempt to save time. Wotan also conceded more if it thought the total value would be increased by the offer. Wotan used neutral language and emotions.

1.4.5 Agent Murphy

Agent Murphy utilized several innovations, including a preference graph which was updated after user statements. It also took a pessimistic view of the human's preferences, tending to assume the worst possible outcome. Murphy also attempted to lighten the mood with occasional jokes.

1.4.6 LyingAgent

LyingAgent focused on perpetrating a type of lie often referred to as the "fixed-pie lie". By misleading the human player into believing they were seeking the same items, it was able to concede items that seemed valuable, but were actually nearly worthless to it. This, plus its generally friendly demeanor, allowed it to appear fair while claiming a greater share of value than its human opponent. The LyingAgent has been described in a number of subsequent studies by the authors of that agent following the competition [17].

2 COMPETITION DESIGN

2.1 IAGO Negotiation Platform

The IAGO Negotiation platform was proposed and designed by Mell et al. and was selected to be used for the Human-Agent League of ANAC [12]. The IAGO platform provides several key aspects that make it suitable for running a competition. Firstly, IAGO provides a front-facing GUI for the human-participants (see Figure 1). This GUI is web-based, and does not require plugins beyond JavaScript support. As such, it is supported on multiple browsers and systems, and can be utilized over the web without lengthy installation or training. In particular, this feature allows subjects to be recruited using online experimentation/task platforms, such as Amazon's Mechanical Turk (MTurk).

Secondly, IAGO provides a number of the features necessary for simulating the characteristics of human negotiation. These include an expanded set of channels for communication between the two sides of negotiation, such as by sending text, expressing preferences, transmitting emotions (by sending "emojis"). This is in addition to the traditional methods supported by agent-agent negotiation platforms, such as exchanging offers. IAGO also allows offers to be sent that do not involve all the items in the negotiation ("partial offers"), and includes API functionality for introducing artificial delay into an agent's responses. This latter function is particularly relevant for agents that try to simulate more human-like behavior, by adding "thinking time" to some of the offers. Indeed, IAGO exposes an event purely for this purpose—the OFFER_IN_PROGRESS event. Agents choosing to utilize this event can simulate their preparation for an offer by signaling to the user that they are nearly ready, causing the IAGO GUI to show an indicator of a flashing ellipsis.

These features of IAGO mean that it is capable of providing a platform to address the basic features that intelligent negotiating agents require. It provides information that allows for robust user modeling, and allows multiple channels for communicating in different ways—potentially, against different types or personalities of users, as detected on-the-fly. IAGO provides information that agents require to reason about their own preferences, and allows them to pursue a number of more complex strategies that require specific features (such as partial offers).

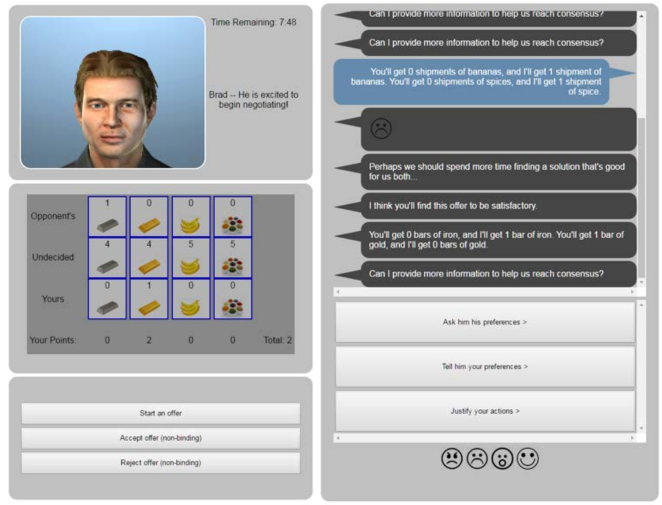

Figure 1: IAGO Research Platform (Client View)

2.2 Human-Agent Competition Design

2.2.1 General Information

The competition featured an array of participant-submitted agents competing against humans in a single, 10-minute multi-issue negotiation. Prior to, and at the end of the negotiation, participants were asked a series of questions, ranging from demographic information to reviews of the agent behavior. The submitted agents were judged along two categories, and prizes were awarded to the best two agents in each category. The first category, agent points, was purely performance-based—agents that received more points in the negotiation were scored higher (regardless of the points the human negotiator scored). The second category, agent likeability, was determined by user-submitted responses to Likert-scale questions after the end of the negotiation. See Section 3.1 for more details on this measure.

2.2.2 Developer-Provided Information

Previous experiments using the IAGO negotiation platform, as well as existing research into the multi-issue bargaining task (a "gold standard" task for negotiation), guided the design of the competition task itself. Agent designers were provided with a set of guidelines that restricted the domain of the negotiation within moderate bounds, but they were not given the details of the task itself, which was determined secretly prior to agent submission. Specifically, the task was publically defined to contain "no fewer than 3 distinct issues, and no greater than 5. Each issue [has] fewer than 20 items." This guideline was further narrowed to define the utilities of both sides per the following equation:

$$\sum_{i=1}^{k} \text{Agent\_utility}(i) * (\text{num\_levels}(i) - 1) = \sum_{i=1}^{k} \text{Human\_utility}(i) * (\text{num\_levels}(i) - 1)$$

where k is the total number of issues. Succinctly, this means that the total for each side would be the same if that side got every item.[2] In this way, the agent designers were given some idea of the scope of the negotiation.

Designers were also provided with a limited set of natural language utterances that the humans could use in the negotiation (Table 1). The human players could express these phrases by selecting them from a pre-created menu. Human players could also send messages that contained information about their preferences, in additional to using the emotional and offer channels. Agents were unrestricted in the types of messages they could send back to players.

Table 1: Available Utterances

It is important that we both are happy with an agreement.
I gave a little here; you give a little next time.
We should try to split things evenly.
We should each get our most valuable item.
Accept this or there will be consequences.
Your offer sucks.
This is the last offer. Take it or leave it.
This is the very best offer possible.
I can't go any lower than this.
We should try harder to find a deal that benefits us both.
There's hardly any time left to negotiate!

Finally, agent developers were provided with the source code for a baseline agent ("Pinocchio") which was provided with the IAGO platform. Details about this agent's features are discussed by the IAGO platform's creators in previous work [11].

2.2.3 Full Negotiation Specification

In the negotiation that was specified for the competition, a single, 10-minute negotiation was conducted. This negotiation had 4 issues, with a varying amount of levels to each. Respectively, each issue had 3, 2, 6, and 3 items. The task was partially integrative, with both sides gaining the most points by receiving the 6-item issue, but differing on their preferences for the two 3-item issues. Both sides also included a BATNA (Best Alternative to Negotiated Agreement), which gave both players a minimum number of points should they fail to reach agreement. In this case, it was 6 points for both sides.

[2] Taken from the ANAC Human-Agent League Competition Guidelines, 2017

2.2.4 Participant Information

Competition subject participants were selected from the MTurk subject pool. Subjects were be adults in the US (18 years or older), and asserted that they were permanent residents of the US (verified with IP address tracking). Restriction to the US was chosen in order to reduce cross-cultural variance. Each submitted competition agent was tested against 25 participants. Participants were not re-used or matched against more than one agent. 54.7% of the participants were male, although no attempt to balance gender across submitted agents was made. Due to the fact that MTurk participants were US-restricted and natural language statements are used in the utterance set of the competition, participants were also asked to affirm that their first language was English. Subject basic demographic information was collected, and they were asked a set of verification questions/attention checks to ensure they comprehended and were engaged in the negotiation, and those that failed were removed from analysis.

All participants were presented with a tutorial of the system before use. Participants were paid near the average hourly rate on MTurk regardless of their success in the negotiation. However, they were also awarded "lottery tickets" based on their performance. These lottery tickets then entered them into a prize drawing for one of several $10 MTurk credits, incentivizing good performance during the negotiation. Subjects were removed from the pool if the agent against which they were matched encountered an unrecoverable crash. In this case, the subjects were not rerun. 10 total subjects were removed in this fashion, 9 of them for the Murphy Agent (and 1 from Agent Wotan). This design allowed the competition to follow best practices for subject recruitment and handling, in line with other research [1].

3 RESULTS & ANALYSES

3.1 Method

For the purposes of the competition, all agents' score and likeability rating (see below) averages were compared one-to-one. Dunnett's 2-sided test confirmed any significant differences for one-way contrasts against the baseline agent, Pinocchio. Significant differences between submitted agents were determined with post-hoc analysis, with no significant differences between Bonferroni correction and Tukey's HSD. For winners, selection of first-place vs. runner up was broken in the direction of the statistical trend for prize-awarding purposes, if no significant difference was observed.

Likeability was determined by a series of self-reported questions that were answered by the participants after the negotiation. These questions referred to how much the human participant liked their partner, and if they would rate the agent highly on several personality measures (such as being "cooperative" or "friendly"). Likeability was determined by the following 7-point Likert questions:

- How satisfied were you with the final agreement?
- How much do you like your opponent?
- Would you negotiate with this opponent again?

Cronbach's alpha for these questions was 0.880, which indicates high reliability. Although not all individual questions varied significantly across agents, none opposed the statistical trend of the composite Likeability measure.

Behavioral measures of the agents and human players are summarized at the end of this section in Table 2, and are discussed in detail in Sections 3.4 and 3.5.

3.2 Likeability

All agents were less likeable than the baseline agent, although the difference between Agent Wotan and the baseline was only marginally significant ($p = .068$). Specifically, the baseline had a mean likeability rating that was 1.18 points higher than Wotan. Agent Wotan and Elphaba were the two most likeable, after the baseline (Figure 2). Agent Wotan and Elphaba were also not significantly different from each other, with a mean likeability difference of only .287 ($p = 1.00$). However, Elphaba was significantly worse than the baseline (differing by 1.47, $p = .010$).

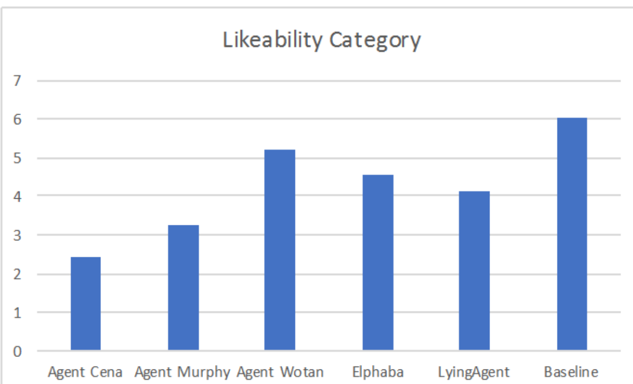

Figure 2: Agent Likeability (7-Point Likert)

3.3 Agent Score

Agent score took into account the total agent points earned in the negotiation. It did not take into account any measures of "social good" or "integrative potential" such as the Nash Product (product of the points of each participant) or Joint Value (sum of the points of each participant). Here, the top performing agent was the LyingAgent. Agent Cena was the runner-up in terms of agent point score, and both it and the LyingAgent scored significantly more than the baseline per ANOVAs with post-hoc Bonferroni corrections. Agent Cena featured a mean difference of 6.86 points over baseline ($p < .01$), and a mean difference of 2.78 points under the LyingAgent ($p > .05$). Similarly, the LyingAgent beat the baseline with a mean difference of 9.64 points ($p < .001$). As such, the difference between Agent Cena and the LyingAgent was not significant. These scores are summarized in Figure 3.

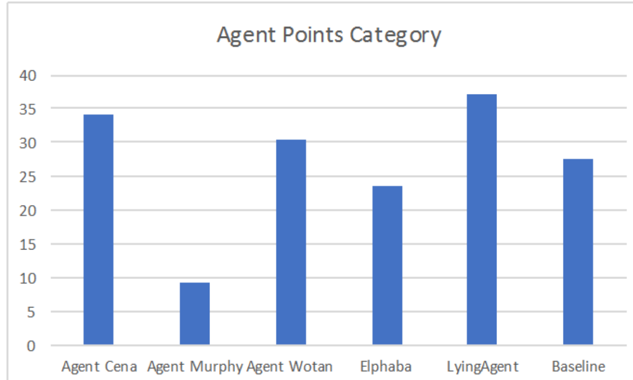

Figure 3: Agent Points (Averages)

3.4 Agent Behaviors' Effects on Score

Since intelligent negotiating agents require specific features (such as user modeling) in order to function, the analyses can offer some insights on the different ways in which these features are implemented in actual negotiation. What makes negotiation somewhat different from other, more restricted types of studies (such as user-behavior, psychologically-driven studies), is that it is not always clear what the actual drivers of certain observed behaviors are. If an agent is known to send more offers than another agent, for example, this behavior could be due to the nature of the agent itself, the nature of the human partner, or a feature of the interaction. Agents that predominately make counter-offers (rather than formulating offers unprompted), may make more offers than those that don't, but will do so especially when partnered with a tough opponent. In other words, tough opponents may enter a feedback loop with certain agents, leading to an escalation in offer exchange—the effect on eventual performance may not be clear.

Still, one goal of competitions such as ANAC is to try to determine the specific antecedents to good outcomes when possible—can we determine what sort of behaviors from the agents do lead to increased points, for example? Although Table 2 can show clearly that agents took different tactics (such as using different types of emotions), it is not clear which of these tactics (if any), are universally helpful. Happily, post-hoc regression analysis of some of the more salient factors leads to takeaways for agent design.

The number of offers that the agents make varies largely between agents. And while more offers mean more information exchanged, such a torrent of offers may reveal more about how disagreeable the human partner was, especially if the agents often counter-offer. And indeed, this is the case, although the effects differed even when within the same category of agents (i.e., the "Likeability winners" vs. the "Point winners"). Specifically, the LyingAgent tended to receive more points when it made fewer offers ($t = -3.378$, $N = 21$, $p=.003$). Elphaba also received more points when it made fewer offers ($t = -2.143$, $N = 25$, $p = .043$). Conversely, Agent Cena received more points as it made more offers ($t = 2.354$, $N=25$, $p = .027$). The other agents did not present significant relationships.

These results can be further examined by combining them with the amount of information the agent had available to it. All the agents attempted to model user preferences—indeed Agent Murphy and Elphaba had explicit measures of reliability. Human participants, however, tend to make few offers themselves, so agents often must rely on explicit statements of preferences by the human users themselves. If these preferences are interpreted well by agents, then they should be able to make use of this updated user model to make better offers. We can find a simulacrum of user model accuracy from the data by examining the number of (truthful) statements that users made about their own preferences. Sadly, the majority of the submitted agents did not adequately utilize this information. No significant correlations were found between the number of user preference statements made and the agent's point outcome, except for Elphaba's case, where the result was distinctly *negative* (t = -2.222, N = 25, p = .036).

Nevertheless, information is only as good as it is used. If an agent has a good user model (due to having preference information), then it might only be useful if the agent actually uses this information to make strategically informed decisions. Specifically, the LyingAgent required a very accurate user model in able to successfully perpetrate its lie and claim value. When examining the relationship between the number of offers made by the agent and the amount of user information is gathered, the LyingAgent shows a significant gain. Explicitly, there exists a two-way interaction between the number of offers made by the agent and the number of preference statements received from the user for the LyingAgent.[3] Specifically, when the LyingAgent makes many offers, it performs better if it ALSO has access to the user's preferences, while it performs much worse if it does not. This interaction is detailed in Figure 4. There is no statistical difference in its performance when it does not make many offers. Specifically, there was a main, negative effect of the number of offers made by the agent, controlling for the number of preferences expressed by the human (t = -3.792, N = 21, p = .001). There was a main, positive effect of the number of preferences, controlling for the number of offers (t = 2.424, N = 21, p = .027). Finally, there was an interaction between offers and preferences (t = 2.172, N = 21, p = .044).

3.5 Agent Behaviors' Effects on Likeability

While the point value of a single negotiation represents a good metric for an agent's success, over time, this may be eclipsed by how much the agent is liked by its opponent. Given the choice, people do not often return to negotiations with people (or agents) that they dislike. Therefore, likeability provides another valid metric for measuring success, and it should come as no surprise that these agent behaviors also have an effect on how likeable an it is perceived to be. For the top-scoring agent in likeability (Agent Wotan), the number of messages sent by the agent was correlated to its likeability (t = 2.564, N=23, p = .018). However, for one of the top-scoring agents in points (Agent Cena), the correlation was significant, and negative (t = -2.125, N=25, p = .045). In short, when Agent Cena sent messages, it reduced the likeability rating, while when Agent Wotan sent messages, the likeability rating increased. There was no significant correlation between messages sent by the other top two agents (Elphaba or LyingAgent) and likeability (Figure 5). These results are perhaps not surprising, since in negotiation, the content of the message being sent is more important than the volume of messages, in many cases. Agent Cena, whose dialogue consisted of aggressive remarks, had a markedly different effect on players than did its more conciliatory cousin, Wotan.

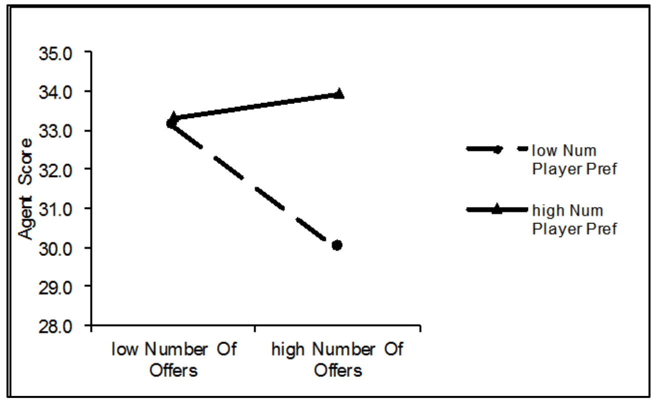

Figure 4: Winning Agent (LyingAgent) – Score by User Preference Statements and Offers Made by Agent

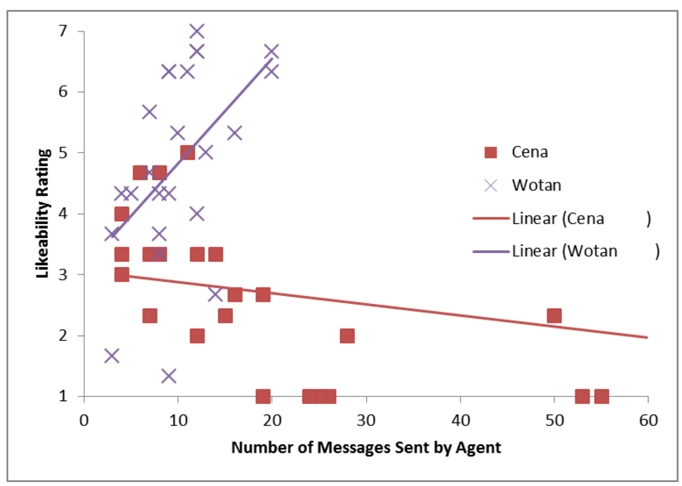

Figure 5: Significant Likeability & Agent Messages Interaction

[3] Note that while the users CAN lie about their preferences, this behavior is tracked in IAGO's data, and is uncommon

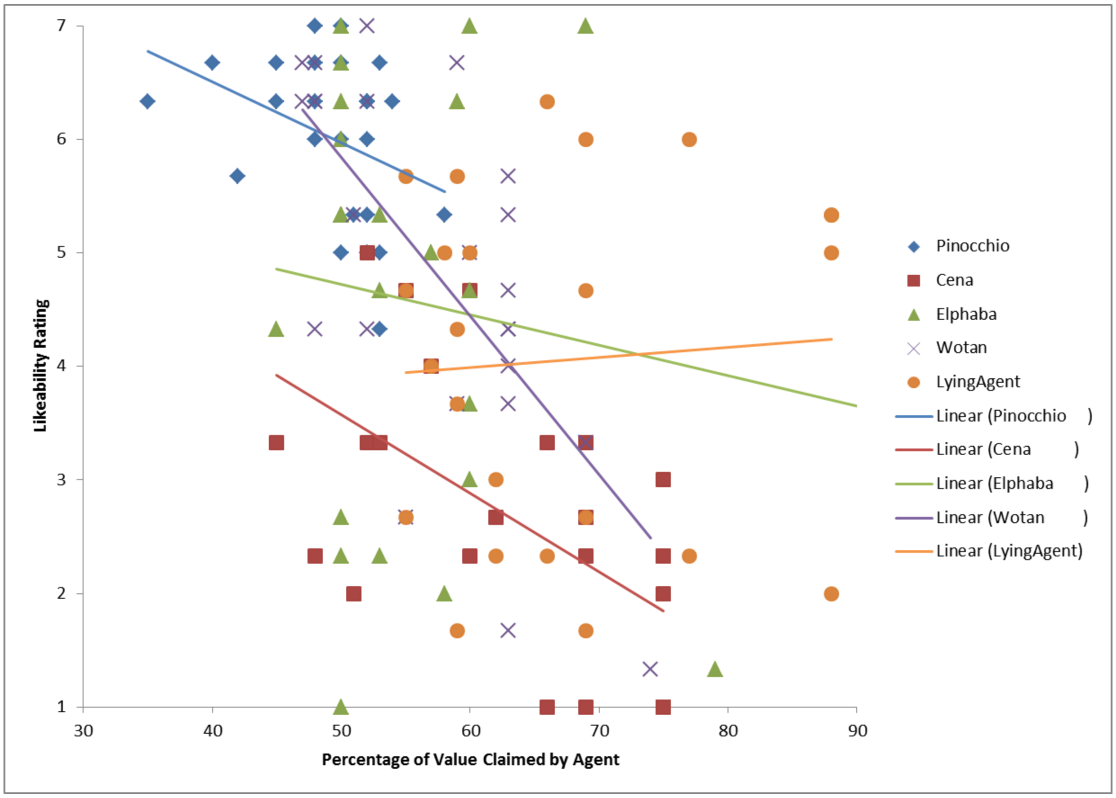

Figure 6: Scatterplot of Niceness Rating vs. Percentage of Value Claimed by Agent (Trend-lines included)

Table 2: Summary Data (Means) for each Agent

	Pinocchio (Baseline)	Cena	Elphaba	Wotan	Murphy	LyingAgent
Likeability	6.01	2.60	4.55	4.83	2.91	4.06
User Points	27.92	19.68	16.96	23.25	8.69	17.92
Agent Points	27.28	34.32	23.56	30.38	9.06	36.88
User Offers	3.20	2.92	5.36	1.21	7.50	2.24
Agent Offers	1.76	2.48	4.52	.71	7.31	1.76
User Preference Statements	.96	.88	1.60	1.04	1.00	2.40
User Non-Preference Messages	4.36	6.40	8.28	3.38	9.75	5.92
User Questions	1.84	1.44	2.36	1.71	2.31	1.68
User Preference Lies	.24	.08	.20	.17	.06	.44
User Happy Emotes	1.20	1.88	1.52	.21	3.00	.68
Agent Happy Emotes	3.08	2.60	1.88	.21	3.94	1.28
User Angry Emotes	.08	4.80	.64	.08	.50	.28
Agent Angry Emotes	.00	8.72	2.16	.00	.00	.00
User Sad Emotes	.00	.72	.44	.04	.38	.24
Agent Sad Emotes	1.64	.00	12.84	.25	.00	2.16
User Surprised Emotes	.28	1.76	.32	.08	.56	.68
Agent Surprised Emotes	.12	.00	.32	.00	1.06	.00

Because human negotiators are often concerned with ideas of fairness, agents that can successfully seem "fair" most likely have a good user model (and good strategic use of it). The final distribution of the points in the negotiation can be examined to determine the actual fairness of a deal (although this does NOT correlate exactly to the perceived fairness of a deal, a fact exploited by the LyingAgent). By examining the percentage of the total points that went to the agent, the relationship between this outcome and likeability can be analyzed. For two of the top-scoring agents (Cena and Wotan), this relationship was significant: ($t = -3.37$, $N = 25$, $p = .003$) for Cena and ($t = -4.34$, $N = 24$, $p < .001$) for Wotan. When either agent took more than its fair share of the points in a negotiation, its likeability suffered. Interestingly, this negative trend existed for all agents, except the LyingAgent, where the trend was slightly positive (but not significant). See Figure 6.

4 DISCUSSION & CONCLUSIONS

The agents submitted to this competition attempted to solve the problems required of intelligent negotiators: they modeled their opponents, determined their own strategies, and adapted to human behavior. In doing this, they followed a number of varied strategies which utilized all the channels supported within IAGO: messaging, offer exchange, and emoting. The two top-performing agents relied on using advanced tactics such as expression of anger (Agent Cena) or lying about the agent's own preferences in order to claim value (LyingAgent). Furthermore, the two most likeable agents (beyond the baseline), also utilized emotion and dialogue to create rapport with the human participant (Agent Wotan & Elphaba).

What is perhaps most striking about these results is how varied the effective tactics seem to be. Agents that send a great deal of messages to the player CAN lead to increased likeability ratings (in the case of Agent Wotan), or reduced ratings (Cena) but this is not universal. Indeed, saying the right things at the right time is quite important—a relatively new dimension within ANAC, where freeform dialogue is normally not allowed between agents.

Moreover, sending offers could be either a blessing or a curse to an agent, depending on how they are implemented. For the blustering, aggressive Agent Cena, a constant barrage of offers increased its performance, whereas unplanned offers did not serve either LyingAgent or Elphaba well. However, armed with knowledge of an opponent's preferences as well, LyingAgent was able to claim value through judicious offers. This knowledge must be applied carefully though, as Elphaba appeared to alter its strategy in the presence of player preferences in a way that actually decreased its performance. In short, automated negotiating agents need to apply the right tactics in the right circumstances. Certainly, the use of anger in negotiation may work well, as can strategic lying, or rapport building techniques.

And while the relationship between scoring well and liking one's opponent is certainly predictable, these results illustrate a very important point about how humans perceive outcomes. The LyingAgent worked by convincing humans that they were getting a fair deal (when they were actually much closer to a 70/30 split). The LyingAgent had a great deal of variance in how it was perceived (see Figure 6), but when it was able to successfully sell its lie, it appeared to do well.

In the future, we hope to reduce the divide between the categories of agents into "likeability" agents and "point-scoring" ones. Future competitions within ANAC in the Human-Agent League will involve repeated negotiations, in which reputational effects may cause the most likeable agents to score well (when the long-term benefits of their strategies may shine). Furthermore, short-sighted strategies such as those employed by Agent Cena and LyingAgent may backfire, with scores plummeting in later rounds. For these reasons and more, we look forward to the continued evolution of negotiating agents, and their performance in future competitions.

ACKNOWLEDGMENTS

This work was partially supported by a variety of grants and organizations, which have been redacted for blind review. The authors of this paper also wish to thank the organizers of ANAC for their support of new leagues within the competition, as well as their provision of the data set. Finally, we wish to thank the researchers who designed and submitted their agents to the competition, without whom this work would not have been possible.

REFERENCES

[1] Alto, K M , McCullough, K M and Levant, R F , 2018 Who is on Craigslist? A novel approach to participant recruitment for masculinities scholarship *Psychology of Men & Masculinity*, 19(2), p 319

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

[3] Baarslag, T , Kaisers, M , Gerding, E , Jonker, C M and Gratch, J , 2017 When will negotiation agents be able to represent us? The challenges and opportunities for autonomous negotiators In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17).*

[4] Broekens, J , Harbers, M , Brinkman, W P , Jonker, C M , Van den Bosch, K and Meyer, J J , 2012, September Virtual reality negotiation training increases negotiation knowledge and skill In *International Conference on Intelligent Virtual Agents* (pp 218-230) Springer, Berlin, Heidelberg

[5] de Melo, C M , Marsella, S and 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

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

[7] Gratch, J , DeVault, D , Lucas, G M and Marsella, S , 2015, August Negotiation as a challenge problem for virtual humans In *International Conference on Intelligent Virtual Agents* (pp 201-215) Springer, Cham

[8] Jonker, C M , Aydogan, R , Baarslag, T , Fujita, K , Ito, T and Hindriks, K V , 2017 Automated Negotiating Agents Competition (ANAC) In *AAAI* (pp 5070-5072)

[9] Leviathan, Y, and Matias, Y, 2018, May Google Duplex: An AI System for Accomplishing Real-World Tasks over the Phone Blog post, downloaded from https://ai googleblog com/2018/05/duplex-ai-system-for-natural-conversation html

[10] Lin, R , Oshrat, Y and Kraus, S , 2009, May Investigating the benefits of automated negotiations in enhancing people's negotiation skills In *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 1* (pp 345-352) International Foundation for Autonomous Agents and Multiagent Systems

[11] Mell, J and 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 IAGO

[12] Mell, J and Gratch, J , 2016, May IAGO: interactive arbitration guide online In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (pp 1510-1512) International Foundation for Autonomous Agents and Multiagent Systems IAGO

[13] Mnookin, R H and Susskind, L E eds , 1999 Negotiating on behalf of others: advice to lawyers, business executives, sports agents, diplomats, politicians, and everybody else (Vol 1) Sage Publications

[14] Peled, N and Kraus, S , 2015 A study of computational and human strategies in revelation games *Autonomous Agents and Multi-Agent Systems*, 29(1), pp 73-97

[15] Reeves, B and Nass, C I , 1996 The media equation: How people treat computers, television, and new media like real people and places *Cambridge university press*

[16] Rosenfeld, A , Zuckerman, I , Segal-Halevi, E , Drein, O and 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

[17] Nazari, Z , Lucas, G and Gratch, J , 2017, August Fixed-pie Lie in Action In *International Conference on Intelligent Virtual Agents* (pp 287-300) Springer, Cham

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

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

[20] Van Kleef, G A , De Dreu, C K and Manstead, A S , 2004 The interpersonal effects of anger and happiness in negotiations *Journal of personality and social psychology*, 86(1), p 57

[21] Accepted, upcoming publication, redacted for anonymous review Upcoming in *Proceedings of the 2018 International Conference on Autonomous Agents and Multiagent Systems* International Foundation for Autonomous Agents and Multiagent Systems