# The Impact of Implicit Information Exchange in Human-agent Negotiations

Emmanuel Johnson and Jonathan Gratch
University of Southern California
Playa Vista, CA USA
{ejohnson, gratch}@ict.usc.edu

## ABSTRACT

Intelligent virtual agents have been developed to study, assess and teach a variety of human interpersonal skills. Here we examine the impact of an agent's perspective-taking sophistication on human negotiators. Good perspective-takers can discover creative solutions that benefit both parties, but many have difficulty with this skill. In particular, novices focus on explicit goal-statements (e.g., "I want apples more than bananas") but discount goal-relevant information implicit in the opponent's offers. Many human-agent negotiation agents similarly ignore implicit information. We examined the influence of implicit information on human negotiators by independently enhancing agents in two ways: do agents *communicate* implicit information and do they *attend* to implicit information communicated by users. We find that communicating implicit information seems to confuse user's perspective-taking ability, yet paradoxically, helps lead them to better outcomes. In contrast, an agent that attends to user's implicit communications shows better perspective-taking but fails to translate this into better outcomes. These results emphasize the challenges associated with implicit information. We discuss how these results impact the design of negotiation agents for applications, analysis and pedagogy.

## CCS CONCEPTS

• Human-centered computing • Collaborative and social computing

## KEYWORDS

Opponent Modeling, Human-agent Negotiation

**ACM Reference format:**

Emmanuel Johnson and Jonathan Gratch. 2020. The Impact of Implicit Information Exchange in Human-agent Negotiations. In *Proceedings of ACM Woodstock conference (AAMAS'20). ACM, New York, NY, USA, 8 pages.*

## 1 Introduction

Intelligent virtual agents have been developed to study [1-3], assess [4] and teach a variety of interpersonal skills [5]. One important skill that has received considerable interest is negotiation. Negotiation serves as a challenge problem for the IVA community [6] – it engages a wide range of mental processes including theory-of-mind, deception and emotion – but it also holds considerable practical importance. Inexpert negotiation skills can lead to lower salaries [7] and increased political and social conflict [8, 9].

A core principle of negotiation is perspective-taking, or what is sometimes called *opponent modeling* [10, 11]. By better understanding the goals of one's opponent, it is often possible to discover creative solutions that benefit both parties. Perspective taking is enhanced when parties share information [12]. For example, if one side states they prefer apples and another side state they prefer bananas, parties can realize they hold different goals. But understanding these differences is not enough. Parties must further realize these goals are complementary and that solutions exist that benefit both sides. For example, joint benefits are maximized if one side takes all the apples and the other takes all the bananas. Thus, effective negotiation requires (1) accurately modeling one's opponent and (2) using these models to propose good solutions.

There are many sources of information relevant to perspective-taking. Within the negotiation literature, a distinction is often made between *explicit* information and *implicit* information. Explicit information is generally defined as verbal statements that directly communicate a party's goals, such as "I like apples more than bananas" [13]. However, other aspects of an opponent's behavior can indirectly reveal their goals. For example, expressions of anger can convey a negotiator has high aspirations [14]. In this paper, we focus on the information revealed by a negotiator's pattern of offers. For example, if a negotiator keeps all the apples and offers bananas, this suggests that bananas hold less value to them.

Studies of human negotiations suggest they have difficulty with information exchange in general [12] and implicit information in particular, although this may reflect a "low context" bias of Western culture [15]. Agents that negotiate with humans often hold a similar bias. For example, IAGO [16] is a popular framework for building agents that negotiate with people and has been used to create dozens of agents as part of the Automated Negotiation Agents Competitions [17]. By default, IAGO fails to attend to implicit information when modeling the goals of human negotiators. Further, it minimizes the implicit information communicated through its offers. Interestingly, much of the research on "rational" agents that negotiate with other agents has shown the opposite bias: such algorithms typically *only* communicate and attend to implicit information, as explicit statements are viewed as "cheap talk" [18].

In this paper, we extend IAGO's sophistication with two abilities related to implicit information. First, we extend IAGO's opponent-modeling algorithm to attend to implicit information. Second, we extend IAGO's offer-generation mechanism to manipulate the


*WOODSTOCK'18, June, 2018, El Paso, Texas USA*

DOI: 10.1145/3383652.3423895

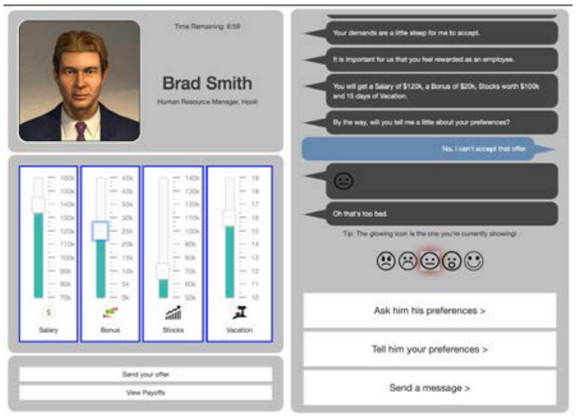


**Figure 1: A salary negotiation implemented in IAGO**

transparency of its offers (i.e., how clearly do the offers reveal the agent's goals). Research on rational agents suggests that outcomes should be improved by both enhancements: (1) humans should be able to form better models of agents when more information is provided (2) agents should form better models of humans by attending to more information, and (3) this improved perspective taking should translate into better negotiated outcomes. Research on actual human behavior suggests a more equivocal outcome and that users may ignore or become confused by implicit information.

We examine these issues through an experiment that independently manipulates each capability. The results highlight the limited utility of rational model in predicting human outcomes. As suggested by psychological findings, we find that communicating implicit information seems to confuse user's perspective-taking ability, yet paradoxically, this information helps lead them to better outcomes (though they were less unsatisfied with these gains). In contrast, agents that attend to user's implicit communications show better perspective-taking as predicted by rational models, but this improved accuracy fails to translate into better user outcomes. These results emphasize the problems associated with implicit information. We discuss how these results impact the design of negotiation agents for applications, negotiation theory and pedagogy.

## 2 Background and Research Questions

**Multi-issue bargaining:** Negotiation can take many forms. We restrict ourselves to a standard abstraction used to study and teach negotiation (referred to as a multi-issue bargaining task with linear additive goals [19]). Here, a pair of negotiators must find a compromise across a range of issues. For example, Figure 1 illustrates a four-issue salary negotiation involving salary, bonus, stocks and vacation. Each issue, $i$, has a discrete number of levels (10 in Figure 1). Players bid by specifying a level for each issue (this defines an *outcome space,* denoted $\Omega$, of all possible outcomes). In Figure 1, there are $10^4$ possible outcomes. The utility of an outcome for a player is expressed by a linear utility function. Each player assigns a private numeric goal-weight, $w_i$, to each issue. The utility of an outcome, $u(\omega)$ depends on the level, $l_i$, assigned to each issue:

$$u(\omega) = \sum_{i=1}^{n} w_i \cdot l_i$$

The nature of the negotiation changes dramatically depending on the issue-weights for each player. If both parties have the same weights, the negotiation is a zero-sum (oft called a "fixed-pie" negotiation). When opponents hold different goals, there are opportunities to create value for both sides by making tradeoffs across issues (e.g. offering more stocks in exchange for less vacation days). As these weights are typically private, perspective-taking may be required to realize these opportunities.

Unfortunately, human negotiators often fail to realize these opportunities. Negotiators are often poor at perspective taking. They often hold a "fixed-pie bias" [20], meaning they assume the negotiation is zero-sum. They also fail to exchange information or miss what information has been shared. Even when they understand their opponent's perspective they may fail to exploit this this understanding. For example, they may fail to appreciate the importance of making tradeoffs to realize win-win solution.

Several measures have been developed to quantify these errors in perspective taking and value-creation. In this article, we used Kendall's rank correlation coefficient as a measure of perspective-taking accuracy and joint points as a measure of value creation. These will be detailed in the experimental section below.

**IAGO Extensions:** The AI community has developed a number of algorithms to play the multi-issue bargaining task, either with other agents (e.g., [18]) or with humans [21-23]. In this paper, we explore extensions to the popular IAGO framework [16], and specifically the "Pinocchio" agent that is provided by default [24]. Human players can exchange information with this agent. For example, they can reveal the pairwise ranking of their priorities (I like X more than Y) or request the same (Do you like X more than Y?). Pinocchio implements perspective-taking based on this information (it develops a model of the human opponent from these explicit preference statements). The more truthful information an opponent reveals, the more accurate these models will be.

Pinocchio plays in a human-like way, which means it implements common biases found in inexperienced negotiators. It holds a fixed-pie bias (it assumes zero-sum unless its opponent reveals information that contradicts this assumption). Like many human players, it only attends to explicit information (it ignores information in the pattern of its opponent's offers). It is a lazy information exchanger: it does not actively reveal its goals but honestly responds when asked. Finally, like many human players, it tries to be fair [25]. It makes fair and efficient offers given its current understanding of the opponent's interests. These design choices make it ideal for examining a human player's skill in creating value with novice opponents (e.g., they fail to create value unless they actively engage in information exchange and perspective taking [24]).

We extend IAGO's sophistication to utilize implicit information in two ways. First, we extend its opponent modeling methods to attend to implicit as well as explicit information. Specifically, we incorporate an opponent modeling method by Nazari that has been shown to more accurately model human goals [11]. Second, we extend its offer-generation method to implicitly convey the agent's goals. By default, Pinocchio makes fair offers that maximize joint outcomes (given its current beliefs about the human's goals). However, in a sufficiently complex negotiation, many offers satisfy this property. A maximally uninformative offer would be the offer from this set that reveals the least information about the agent's goals (e.g., a bid that offered the same level on each issue would reveal no information about what the agent wants). A maximally informative offer would be the offer from this set that conveys the most information (e.g., If the agent values salary the most and vacation the least, an informative bid would offer the greatest concessions on vacation and the least concessions on salary). Note that this extension does not alter the value of offers from the agent's perspective, but can alter their value to the human opponent if the agent has incorrectly modeled the human's goals (this may have complex implications for value creation, which we study below).

**Impact of implicit information on perspective taking:** How might the use of implicit preference information impact perspective-taking ability? From a rational information theory perspective, more (truthful) information should be better. The IAGO agent should develop more accurate models of the human to the extent it attends to implicit information. The human should develop more accurate models of the agent to the extent the agent provides this information. But will these findings hold in practice?

*Will the human be more accurate?* Some research suggests that novice human negotiators will fail to appreciate or even become confused by implicit information. For example, Adair and colleagues studied cross-cultural contexts where negotiators differed in their use of implicit information [15]. Low-context cultures [26], like those studied in this paper, emphasize direct information exchange, whereas high-context cultures utilize indirect information. When these cultures mix, perspective taking and joint outcomes suffer. This effect seems driven by dynamic patterns of information exchange: exchange of direct information promotes accuracy but accuracy falls when direct information leads to indirect responses. In a sense, the sophisticated version of IAGO simulates this cultural mis-match and could undermine perspective taking.

However, other research suggests that human perspective taking is enhanced by interacting with agents with sophisticated perspective-taking skills. For example, Yoshida and colleagues found that people could recognize the degree of recursive reasoning performed by a computer opponent in a version of the Stag Hunt game and adjusted their behavior accordingly [27]. In the realm of negotiation, de Weerd and colleagues showed that participants that played Colored Trails [28] – a testbed to study multi-issue negotiations – exhibited more second-order theory of mind reasoning when they played with cognitively sophisticated agents [29]. Similarly, Stevens and colleagues showed that students that practiced with an ACT-R agent possessing theory-of-mind were able to achieve better outcomes in a single-issue negotiation task [30]. These conflicting perspectives lead to the following research questions:

- Is human accuracy enhanced when agents provide implicit information about their goals?
- Is human accuracy enhanced when agents that attend to implicit information? (as this may stimulate human-perspective taking)
- How do these mechanisms interact to shape human accuracy?

*Will the agent be more accurate?* Prior research suggests that agent perspective-taking is enhanced by attending to the information implicit in the human's pattern of offers. Nazari and colleagues examined the accuracy of different opponent modeling techniques at predicting human goals in human-human negotiations and found the most accurate models utilized both implicit and explicit information [11]. However, these models adopted a third-party perspective (they analyzed traces from completed negotiations). Would results be different if the models were used to alter agent behavior *during* the negotiation? Negotiations are a back-and-forth between parties and the information provided emerges from this complex dynamic. Could early mistakes cause downstream confusion? For example, research on implicit opponent models has found that these techniques are sometimes confused as opponents offers might reflect a compromise between what they want and what they think their opponent wants (consider the fable The Gift of the Magi) [31]. This leads to the following research questions:

- Will agent accuracy increase when agents provide implicit information?
- Will agent accuracy increase when agents attend to implicit information?
- How do these mechanisms interact to shape agent accuracy?

**Impact of implicit information on value creation:** Despite the emphasis given to opponent-modeling in negotiation research, accurate models are neither necessary nor sufficient to discover win-win solutions. Generally speaking, better perspective taking should translate into higher joint gains. However, accurate models can be insufficient for a number of reasons. Even if both parties hold accurate models, they may fail to understand how to use this information effectively. In particular, they may fail to understand the importance of making tradeoffs to realize the win-win potential. If only one party holds an accurate model, the inaccurate party may view win-win offers with suspicion and reject them (see [32]).

In contrast, some research suggests that parties can create value even in the absence of good perspective taking [33, 34]. For example, Pruitt's trial-and-error heuristic suggests if negotiators take turns making "logrolling offers" (these are offers in which a negotiator claims more of what they want and concedes what they don't want [35]), simply focusing on what is "on the table" rather than guessing what their opponent might want, it is possible to effectively create value. Note that IAGO's informative offers, by definition, involve logrolling (the agent keeps more of what it wants and offers what it wants least). Thus, an agent that uses implicit communication could improve joint outcomes though this mechanism. This leads to the following research questions:

- Will joint gains rise when agents provide implicit information?
- Will joint gains rise when agents attend to implicit information?
- How do these mechanisms interact to shape joint gains?
- How do these mechanisms influence satisfaction in the outcome?

## 2 Experimental Design

280 participants were recruited for a 2x2 experimental approved by our institution's ethics board. Participants engaged in a salary negotiation with a computer agent. We independently manipulated the two enhancements to IAGO related to implicit information: (1) the agent either inferred opponent models from the participant's explicit statements alone (unsophisticated perspective-taking) or from both their statements and pattern of offers (sophisticated perspective-taking); and (2) the agent either communicated its goals using explicit statements alone (unsophisticated communication) or using both explicit statements and offer behavior (sophisticated communication). Participants were randomly assigned across the four resulting conditions. All participants were English speakers from the U.S. were recruited via Amazon Mechanical Turk. Of the 280, 48 were removed for failing to pass the attention check questions,

leaving 232 participants (63.9% male; mean age = 38.7). Do to confusion on how to return to the survey software after the negotiation, an additional 42 participants failed to complete some or all the post negotiation survey. Thus, analysis of participant accuracy involves a reduced sample of 190 participants.

### 2.1 Negotiation Task

Participants engaged in a 4-issue salary negotiation (salary, bonuses, stock options and vacation days). They assumed the role of an employee seeking a position at a large technology company and the agent assumed the role of an HR manager. For each issue, participants and the agent had to agree upon one of ten levels. For example, they could negotiate a salary between $70,000-$160,000 in $10K increments (see Figure 1). Each party received points based on the level they negotiated for each issue (see Table 1). For example, if participants negotiated a salary of $90,000 ($l$=3), they would receive 15 points. The total number of points they received was summed across the four issues. Neither the agent nor the participant knew the other’s preference. In addition to these goals, participants were told they would receive only six points (their BATNA) if they failed to reach an agreement within the seven minutes allotted.

**Table 1: Participant and Agent’s Payoff Matrix**

| | Salary | Bonuses | Stocks | Vacation |
|---|---|---|---|---|
| Agent | 5*(10-*l*) | 1*(10-*l*) | 2*(10-*l*) | 3*(10-*l*) |
| Human | 5**l* | 3**l* | 2**l* | 1**l* |

Although salary and stock options are zero-sum issues, the task affords opportunity to find win-win solutions. Value can be created by trading off between bonuses and vacation days. The joint value of the final deal is maximized if the participant claims more bonuses and fewer vacation days. Thus, there is potential to “grow the pie” if participants correctly understand the other side’s perspective (i.e., accurately model their opponent’s goals).

To motivate performance, participants received a fixed participation fee (approximating the US Federal minimum wage) plus an incentive based on performance: they received lottery tickets proportional to the points they earned in the negotiation, and these were entered into a $100 USD lottery. When presented the instructions, participants where quizzed on several aspects of the system and attention checks were used to filter inattentive participants.

### 2.2 Intelligent Agent Behavior

All agents were developed using the IAGO agent framework. The interface, seen in Figure 1, has a simple virtual character that can convey facial expressions. A text box summarizes the history of the negotiation. Participants can communicate with the agent via menus. These allow the participant to ask questions about what the agent wants, offer information about what they want, and exchange offers. Participants can also convey emojis and send some canned messages (e.g., “This is my final offer. Take it or leave it”). IAGO provides libraries for many of the core functions required to negotiate, such as techniques for opponent modeling and offer generation. To create the experimental factors we modified the default “Pinocchio” agent (see [24] for details of this agent).

*2.2.1 Agent Perspective-Taking Sophistication:* To manipulate perspective-taking sophistication we replaced the default modeling behavior of Pinocchio with a set of “frequency models” proposed by Nazari [36]. Frequency models are a heuristic method for learning what an opponent wants by counting the frequency of certain events, such as how often someone makes a concession on an issue or how often they explicitly assert a preference for that issue (see an overview in [19]). These models estimate a weight for each issue, which can be used to infer the relative importance of each issue. Nazari proposed a hybrid model that models implicit and explicit information separately and averages the resulting weights to combine this information. For implicit information, the model examines offer concession. If an issue ($i$) is claimed in an offer ($k$), ($l_k$ ) indicates how much of that issue was claimed for a negotiator and how much level was assigned to their opponent ($l'_k$ ). To calculate the weight for each item, the implicit model computes a ratio of the items claimed for self, divided by items given to opponent.

$$w_i = \frac{l_k}{l'_k}$$

The explicit model estimates weights from the explicit preference statements by counting how often a positive or negative statement is made about an item. For example, "I like gold more than spices" is a positive statement about gold whereas "I like gold less than bananas" is a negative statement towards gold. The weight for each item $i$ is computed as follows:

$$w_i = |P_i| - |N_i|$$

where Pi is a count of all positive statements made about an item *and* $N_i$ is a set of all negative statements made about an item.

We made one change to align these models with IAGO’s fixed-pie assumption. When weights are equal between two issues (meaning the model is indifferent on their ranking), we assume the ranking is consistent with the agent’s own goals (e.g., if the agent prefers stocks to vacation and the model returns equal weight for these issues, the agent assumes the human prefers stocks to vacation).

Sophisticated-modeling agents used the combined model, whereas unsophisticated agents used the explicit-statement model.

*2.2.2 Agent Implicit Communication:* To manipulate how the agent communicates its goals, we altered the agent’s offer generation model. By default, Pinocchio tries to generate efficient and approximately fair offers. Given its current estimate of the participant’s goals (as estimated in Section 2.2.1) it calculates the Pareto Frontier and generates an offer claiming 60% of the total possible value. However, in a sufficiently complex negotiation problem (like the one in this study), there will be a large number of offers that satisfy this criterion. We introduce an additional constraint to manipulate communicative-sophistication.

The agent with “unsophisticated communication” selects an offer from this set that minimizes how much is revealed about its issue-weights. Specifically, it makes an offer that minimizes the variance in the levels across the issues under negotiation. For example, the offers 5,6,6,5 and 4,9,7,6 both yield the agent the same number of points (60% of the pie, assuming the agent has perfect knowledge of what the participant wants), yet the former has less variance, and therefore provides fewer clues about what the agent truly wants. In contrast, the latter offer suggests the agent cares the most about salary (issue 1), the least about bonuses (issue 2), and so forth.

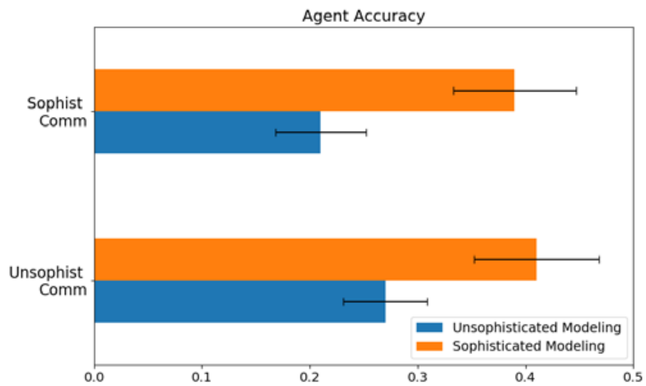


**Figure 2: Agent perspective-taking accuracy (τ)**

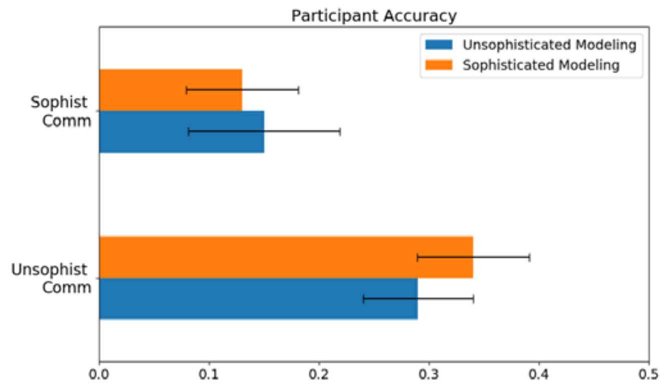


**Figure 3: Participant perspective-taking accuracy (τ)**

The communicatively-sophisticated agent makes offers that maximally conveys its own interests. For example, it would concede the most on vacation, then stocks, then bonuses, then salary.

*2.2.3 Impact of sophistication on agent offers:* Both forms of cognitive sophistication (opponent modeling and communicative) will change how the agent negotiates and this could impact outcomes. As the agent is trying to make efficient offers, its offers change as it forms a better model of the opponent's goals. For example, as the agent learns that the human prefers bonuses to vacation, it will begin to make logrolling tradeoffs across these issues. In theory, this should convey to the participant that the agent is attentive and responsive to their interests. If the human never explicitly asserts their goals, the unsophisticated agent will fail to show this responsiveness, yet the sophisticated agent still will, as it also attends to the human offers. Note also that the actual efficiency of offers depends on the accuracy of the opponent model so that better models should translate into the participant receiving better offers.

Similarly, the communicative sophistication could impact the value of offers the human receives. If the agent's opponent model is correct, offers should have equivalent value, regardless of the communication strategy (even if the levels across issues differ). But if the opponent model is wrong, the communication strategy will impact the value that the human receives. In that the human's counteroffers are likely shaped by the agent's last offer, this can have complex interactions on the final deal. Thus, when humans perform better against sophisticated agents, it is important to disentangle how much this improvement results from better perspective-taking on the part of the human or arises despite the human.

### 2.3 Measures

**Perspective-taking Accuracy:** To measure perspective-taking accuracy, we adopt *Kendall's rank correlation coefficient*, also called Kendall's τ coefficient. This measures the accuracy in the relative ranking of issues rather than the exact weight assigned to each issue. The coefficient ranges from +1, denoting the correct ranking to -1, denoting the reverse ranking. Rank-based measures are particularly relevant as IAGO only allows users to express rankings through their explicit information exchange. After a negotiation, players are asked to rank the importance their opponent assigns to each issue and this is compared with the agent's true goals. For the agent, the final agent model is compared with the true human goals.

**Value Creation and Claiming:** We use individual points obtained by each side as a measure of value claiming (how big a slice of pie did they grab). We use joint points (i.e., the sum of individual points obtained by the participant and the agent) as a measure of value creation (how much did the agents grow the size of the pie). Since the Pinocchio agent tries to be fair, individual points should mirror joint gains as the agent attempts to split the pie.

**Subjective Impressions:** We asked several post-negotiation questions to index participant's perceptions. Participants were askes two questions about perceived perspective-taking accuracy ("I clearly understood the HR manager's goal" and "The HR manager took my interests into account when making offers") and two questions about the agent's communicative behavior ("The HR manager's verbal statements seemed realistic" and "The HR manager's sequence of offers seemed realistic"). Finally, they were asked if they were satisfied with their final offer. Items were on a 7-point Likert scale from "1-strongly disagree" to "7-strongly agree".

## 3 Results

### 3.1 Agent Perspective-taking Accuracy

We performed a two-way ANOVA examining how agent modeling accuracy was shaped by the experimental factors (see Figure 2). There was only a significant main effect of the sophistication of the opponent modeling approach (F (1,229) = 11.224, p =.001). The agent was more accurate when it inferred the participant's goals from both their explicit statements and their pattern of offers. Accuracy was reduced when the agent used sophisticated communication (i.e., attended to but the participant's statements and their pattern of offers) but this was not significant (F(1,229) = .305, p = .581), nor was there an interaction between the factors (F(2,229) = .010, p = .921).

These results reinforce Nazari's finding that opponent modeling accuracy increases by combining implicit and explicit information.

### 3.2 Human Perspective-taking Accuracy

We performed a two-way ANOVA examining how participant modeling accuracy was shaped by the two experimental factors (see Figure 3). We find that participants were confused when the agent provided implicit information about its goals. There was a main effect of communication type with users being less accurate with communicatively sophisticated agents (F(1,187) = 1.458, p = .002).

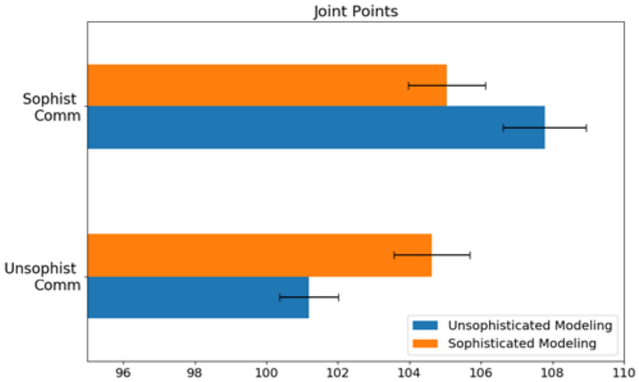


**Figure 4: Value creation (joint points)**

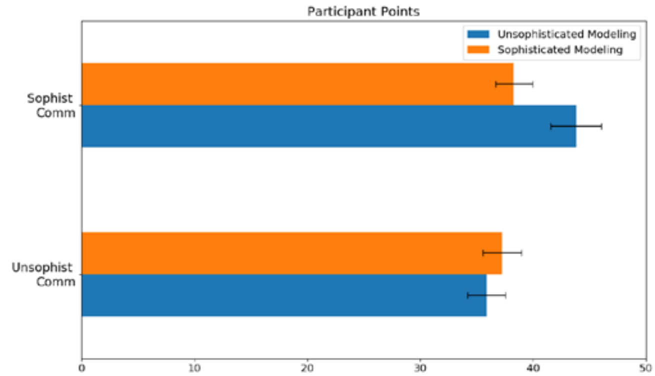


**Figure 5: Value claiming (participant points)**

Participants appeared most confused when the agent used sophisticated modeling, but this apparent difference was not significant (F(1,187) = .74, p = .785), nor were there significant interactions between the factors (F(2,187) = .494), p = .488).

These results are inconsistent with the findings of Yoshida [27]. de Weerd [28] and Stevens [30] – that human perspective-taking can be enhanced through mere exposure to a sophisticated agent – but are consistent with the findings of Adair and colleagues that U.S. negotiators will be confused by implicit information [15].

### 3.3 Value Creation Results

We next used a two-way ANOVA to examine how the experimental factors shaped value creation. (see Figure 4). We find a main effect of communication-sophistication (F(1,230) = 18.275, p<.001). The parties achieved greater joint points when the agent conveyed its goals through its pattern of offers (M=106.2, SD=8.03 vs. M=102.26, SD=6.62). This is interesting given that humans showed the worst perspective-taking in this case, suggesting that human-accuracy was unnecessary to achieve joint gains.

Surprisingly, given the accuracy of agent models, we fail to see a significant effect of modeling-sophistication on joint points (F(1,230) = 0.032, p=.859). We did, however, find an interaction between informative offers and agent modeling technique (F(1,230)=7.260, p=.008). As seen in Figure 4, when the agent made uninformative offers, more sophisticated opponent modeling led to a significant improvement in value creation (F(1,230)=4.090, p=.044 in post-hoc test). However, there was a nonsignificant trend that modeling hurt efficiency when the agent made informative offers (F(1,230)=3.194, p=.075). Note that human users were the least accurate when the agent both attended to and communicated implicit information, perhaps undermining the pair’s ability to create value. The greatest value creation resulted from agents that telegraphed their goals through their offers, but ignored the offer information telegraphed by the human opponent.

Overall, the results are most consistent with the view of Kelly and Pruitt that value can be claimed, despite poor perspective-taking, as long as parties utilize logrolling in their pattern of offers.

### 3.4 Value Claiming

We examined how the experimental factors impact the individual earnings participants receive. As Pinocchio aims for fair deals, we would expect participant points to closely mirror the pattern of joint points. And this is what we see. We observe a main effect of sophisticated communication (F(1,230)=7.546, p<.006) – participants obtained better deals when the agent telegraphed its interests through is offers (M=40.79, SD=14.77 vs. M=36.25, SD=11.29). We also find a trend for an interaction between informative offers and agent modeling technique (F(1, 230) = 3.682, p=.056) with the same pattern as in joint points. Using post hoc tests, we find that when the agent used sophisticated communication, the use of sophisticated modeling slightly undermined participant points. There was no impact of modeling when the agent only communicated with explicit information.

It is possible that participants performed worse in the modeling condition because the agent was able to capture a disproportionate share of the efficiency gains. To test this, we performed a two-way ANOVA on the agent’s points but found no significant differences, ruling out this possibility.

### 3.5 Subjective Impressions

Finally, we performed two-way ANOVAs on the subjective items. The two items indexing perceptions of perspective taking failed to show a difference by condition. Participants somewhat agreed that they understood the agent’s goals (M=5.22, SD=1.284) and that the agent somewhat understood them (M=4.90, SD 1.422). The perceptions of agent behavior showed a significant effect of communicative sophistication. The agents verbal statements were seen as less realistic (M=4.99, SD=1.506 vs. M=5.37, SD=1.095) when the agent provided implicit as well as explicit information (F(1,223) = 8.623, p=.026). Perceptions of pattern of offers showed the same pattern (M=4.52, SD=1.618 v. M=5.01, S=1.378; F(1,223) = 6.149, p=.014). This reinforces the interpretation that participants were confused by implicit information.

For satisfaction with the final outcome, there was no main effect of communication (F(1,223) = 2.224, p=.137) or modeling (F(1,223) = .009, p=.923) but there was a significant cross-over interaction between the two factors (F(1,223) = 5.986, p=.015). Post hoc analysis revealed this was driven by the use of implicit communication in the unsophisticated modeling condition. Participants were least satisfied with their outcome when the agent conveyed its preferences through its offers (M=4.98, SD-1.483) compared with when it only used explicit goal statements (M=5.65, SD=.988). Interestingly, they were least satisfied with the offer that yielded them the most points.

## 4. Discussion

Overall, findings suggest that participants were worse at perspective-taking when agents communicate their goals through their pattern of offers, yet paradoxically, this helped participants create value and obtain better deals for themselves. These findings reinforce and integrate two existing observations from the human negotiation literature. First, novice negotiators have a difficult time integrating implicit and explicit information when modeling their opponent's goals [13, 15]. Our findings demonstrate that the same effect occurs when people negotiate with computer agents. Second, poor perspective-takers can still create value if parties highlight tradeoffs through their sequence of offers [33, 34]. This tactic seems to work equally well in negotiations with IAGO agents.

Although implicit communication helped participants obtain better outcomes, their subjective impressions raised some concerns. Participants obtained the best outcomes when the agent communicated with implicit information, but paradoxically, they were the least satisfied with these offers. This suggest that people might reject such beneficial offers because of confusion about how they were created. This relates to the concept of *procedural justice:* negotiators don't simply care about outcomes but also the process and intentions by which the outcome was produced [37, 38]. One potential lesson is that agents that use implicit information may benefit from explicitly highlighting their use of this information (e.g., "Oh, I see you are conceding vacation days, so does this mean you don't value vacation time?). See the work of Crandall[39] for one suggestion along this direction.

Our findings further illustrate that agents are better perspective-takers when they attend to implicit information. Agents were more accurate in inferring participant goals when they combined information from both the participant's statements and their sequence of offers. This reinforces but also extends the findings of Nazari and colleagues [11]. Nazari showed this accuracy can be achieved by analyzing negotiation traces after the fact. Here we show the models also work *during* the negotiation (where the model is used to inform the agent's within-negotiation behavior). This emphasizes the robustness of this approach.

Unfortunately, better agent perspective-taking did not facilitate human perspective-taking, as claimed by some recent studies [27, 29, 30]. The different agent models had no effect on human perspective-taking ability. Enhanced perspective-taking did help create value, but only when the agent refrained from communicating implicit information. This again reinforces the problematic nature of implicit information, at least for U.S. participants.

The equivocal benefits of opponent modeling echo findings in the agent-agent negotiation literature. For example, despite the theoretical importance of theory-of-mind reasoning, a recent meta-analysis of negotiation agents showed that opponent modeling had limited ability to improve negotiation outcomes and, rather, strategies for generating offers had a far larger impact on outcomes [40].

Our findings also emphasize the importance of negotiation-skills training as well as the potential for virtual agents to assist in skill-development. Participants exhibited many of the same errors with our agents that they exhibit in human-on-human negotiations: they fail to take the perspective of their opponent, they are confused by implicit information, and the failing in perspective taking, failure to attend to implicit information and failure to capitalize on opportunities for value creation. This reinforces the potential benefits of using agents to teach negotiation skills [41, 42].

Several limitations qualify these findings. Participants engaged in a single negotiation and were under some time pressure (they had seven minutes to complete the task). This could have imposed cognitive load and the inclusion of implicit information may have simply been too taxing. Perhaps repeated interactions with the task could lead to a different pattern of results. It should also be noted that our experiments were more nuanced than many of the above-mentioned studies. Prior studies tend to compare sophisticated agents with agent with zero theory-of-mind capability. Here, all agents had some perspective taking ability but they differed by degree. Perhaps cognitive sophistication transference would emerge with a blunter manipulation. Of course, the results may depend on some idiosyncrasy of the negotiation task or agent behavior. Thus, these findings need to be replicated on other domains and with other agent architectures.

## Acknowledgments

Research was sponsored by the Army Research Office and was accomplished under Cooperative Agreement Number W911NF-20-2-0053 and the National Science foundation under grant 1822876. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein

## References

[1] C. de Melo, P. J. Carnevale, and J. Gratch, "Using virtual condederates to research intergroup bias and conflict." *74th Annual Meeting of the Academy of Management*, Philadelphia, PA, 2014.

[2] J. Blascovich, J. Loomis, A. Beall, K. Swinth, C. Hoyt, and J. N. Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology." *Psychological Inquiry,* vol. 13, pp. 103-124, 2002.

[3] S. Marsella, D. Pynadath, and S. Read, "PsychSim: Agent-based modeling of social interactions and influence." *International Conference on Cognitive Modeling*, 2004, pp. 243-248.

[4] A. C. Graesser, Z. Cai, X. Hu, P. W. Foltz, S. Greiff, B.-C. Kuo*, et al.*, "Assessment of Collaborative Problem Solving." *Design Recommendations for Intelligent Tutoring Systems,* vol. 275, 2017.

[5] B. Lok, F. Rick, R. Andrew, J. Kyle, D. Robert, C. Jade*, et al.*, "Applying Virtual Reality in Medical Communication Education: Current Findings and Potential Teaching and Learning Benefits of Immersive Virtual Patients." *Journal of Virtual Reality,* vol. 10, pp. 185-195, 2007.

[6] J. Gratch, D. Devault, G. Lucas, and S. Marsella, "Negotiation as a Challenge Problem for Virtual Humans." *IVA2015,* Delft, The Netherlands, 2015.

[7] A. Leibbrandt and J. A. List, "Do Women Avoid Salary Negotiations? Evidence from a Large-Scale Natural Field Experiment." *Management Science,* vol. 61, pp. 2016-2024, 2015/09/01 2014.

[8] R. L. Hall, "Measuring Legislative Influence." *Legislative Studies Quarterly,* vol. 17, pp. 205-231, 1992.
[9] T. Bosse, C. Gerritsen, and J. de Man, "An intelligent system for aggression de-escalation training." *22nd European Conference on Artificial Intelligence*, 2016, pp. 1805-1811.
[10] T. Baarslag, M. J. C. Hendrikx, K. V. Hindriks, and C. M. Jonker, "Learning about the opponent in automated bilateral negotiation: a comprehensive survey of opponent modeling techniques." *AAMAS2015,* pp. 1-50, 2015.
[11] Z. Nazari, G. Lucas, and J. Gratch, "Opponent Modeling for Virtual Human Negotiators." *IVA2015.* Delft, the Netherlands, 2015.
[12] L. L. Thompson, "Information exchange in negotiation." *Journal of Experimental Social Psychology,* vol. 27, pp. 161-179, 3// 1991.
[13] W. L. Adair, T. Okumura, and J. M. Brett, "Negotiation behavior when cultures collide: the United States and Japan." *Journal of Applied Psychology,* vol. 86, p. 371, 2001.
[14] G. A. van Kleef, C. de Dreu, and A. S. R. Manstead, "The interpersonal effects of anger and happiness in negotiations." *Journal of Personality and Social Psychology,* vol. 86, pp. 57-76, 2004.
[15] W. L. Adair, "Integrative sequences and negotiation outcome in same-and mixed-culture negotiations." *International Journal of Conflict Management,* vol. 14, pp. 273-296, 2003.
[16] J. Mell and J. Gratch, "IAGO: Interactive Arbitration Guide Online." *AAMAS2016*, 2016, pp. 1510-1512.
[17] J. Mell, J. Gratch, T. Baarslag, R. Aydrogan, and C. Jonker, "Results of the First Annual Human-Agent League of the Automated Negotiating Agents Competition." *IVA2018*, Sydney, Australia, 2018.
[18] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge, "Automated Negotiation: Prospects, Methods and Challenges." *International Journal of Group Decision and Negotiation,* vol. 10, 2001.
[19] T. Baarslag, M. Hendrikx, K. Hindriks, and C. Jonker, "Predicting the performance of opponent models in automated negotiation." *EEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)-Volume 02*, 2013, pp. 59-66.
[20] F. Harinck, C. de Dreu, and A. E. M. van Vianen, "The Impact of Conflict Issues on Fixed-Pie Perceptions, Problem Solving, and Integrative Outcomes in Negotiation." *Organizational Behavior and Human Decision Processes,* vol. 81, pp. 329-358, 3// 2000.
[21] L. O. Ouali, N. Sabouret, and C. Rich, "A computational model of power in collaborative negotiation dialogues." *IVA2017*, 2017, pp. 259-272.
[22] A. Rosenfeld, I. Zuckerman, E. Segal-Halevi, O. Drein, and S. Kraus, "NegoChat: a chat-based negotiation agent." *AAMAS2014.* Paris, France, 2014.
[23] R. Zhao, O. J. Romero, and A. Rudnicky, "SOGO: A Social Intelligent Negotiation Dialogue System." *IVA2018.* Sydney, NSW, Australia, 2018.
[24] J. Mell and J. Gratch, "Grumpy and Pinocchio: The effect of language and strategy in human-agent negotiation." *AAMAS2017*, Sao Paulo, Brazil, 2017.
[25] E. Fehr and K. M. Schmidt, "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics,* vol. 114, pp. 817-868, August 1, 1999 1999.
[26] E. T. Hall, *Beyond culture*: Anchor, 1989.
[27] W. Yoshida, R. J. Dolan, and K. J. Friston, "Game theory of mind." *PLoS computational biology,* vol. 4, e1000254, 2008.
[28] Y. Gal, B. J. Grosz, S. Kraus, A. Pfeffer, and S. Shieber, "Colored trails: a formalism for investigating decision-making in strategic environments." *Workshop on reasoning, representation, and learning in computer games*, 2005, pp. 25-30.
[29] H. de Weerd, E. Broers, and R. Verbrugge, "Savvy software agents can encourage the use of second-order theory of mind by negotiators." *CogSci*, 2015.
[30] C. A. Stevens, J. Daamen, E. Gaudrain, T. Renkema, J. D. Top, F. Cnossen*, et al.*, "Using Cognitive Agents to Train Negotiation Skills." *Frontiers in Psychology,* vol. 9, 2018-February-19 2018.
[31] T. Baarslag, M. Hendrikx, K. Hindriks, and C. Jonker, "Measuring the performance of online opponent models in automated bilateral negotiation." *AI 2012: Advances in Artificial Intelligence*, ed: Springer, 2012, pp. 1-14.
[32] M. K. Lee and S. Baykal, "Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division." *ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2017, pp. 1035-1048.
[33] H. H. Kelley and D. P. Schenitzki, "Bargaining." *Experimental Social Psychology. New York: Holt, Rinehart, and Winston,* pp. 298-337, 1972.
[34] D. G. Pruitt and P. J. Carnevale, *Negotiation in social conflict*: Thomson Brooks/Cole Publishing Co, 1993.
[35] L. Thompson, "The influence of experience on negotiation performance." *Journal of Experimental Social Psychology,* vol. 26, pp. 528-544, 1990.
[36] Z. Nazari, G. M. Lucas, and J. Gratch, "Opponent modeling for virtual human negotiators." *IVA2015,* 2015, pp. 39-49.
[37] M. K. Lee, A. Jain, H. J. Cha, S. Ojha, and D. Kusbit, "Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation." *Human-Computer Interaction,* vol. 3, pp. 1-26, 2019.
[38] J. R. Curhan and H. A. Elfenbein, "What do people value when they negotiate? Mapping the domain of subjective value in negotiation." *Journal of Personality & Social Psychology,* vol. 91, pp. 493-512, 2006.
[39] J. W. Crandall, M. Oudah, Tennom, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon*, et al.*, "Cooperating with machines." *Nature Communications,* vol. 9, p. 233, 2018/01/16 2018.
[40] T. Baarslag, A. Dirkzwager, K. V. Hindriks, and C. M. Jonker, "The Significance of Bidding, Accepting and Opponent Modeling in Automated Negotiation."*ECAI*, 2014, pp. 27-32.
[41] E. Johnson, "Using Automated Agents to Teach Negotiation." in *AAAI Conference on Artificial Intelligence*, 2019, pp. 9888-9889.
[42] J. Broekens, M. Harbers, W.-P. Brinkman, C. M. Jonker, K. Van den Bosch, and J.-J. Meyer, "Virtual reality negotiation training increases negotiation knowledge and skill." *IVA* 2012.