

---

# The Promise and Peril of Automated Negotiators

*Jonathan Gratch\**

---

*Innovations in artificial intelligence are enabling a new class of applications that can negotiate with people through chat or spoken language. Developed in close collaboration with behavioral science research, these algorithms can detect, mimic, and leverage human psychology, enabling them to undertake such functions as the detection of common mistakes made by novice negotiators. These algorithms can simulate the cognitive processes that shape human negotiations and make use of these models to influence negotiated outcomes. This article reviews some of the scientific advances enabling this technology and discusses how it is being used to advance negotiation research, teaching, and practice.*

**Keywords:** negotiation, artificial intelligence, automated negotiation, intelligent tutoring

## Introduction

Many avoid negotiations and this aversion has real economic costs. Consumer Reports found that more than a third of Americans refused to bargain over the price of products and services, yet 90 percent of those who negotiated saved money (Consumer Reports 2013). A recent survey by Glassdoor, a popular job recruiting site, found almost 60 percent of American employees simply accepted the first salary they were offered (Glassdoor Team 2016). This, despite research demonstrating that those

---

\*Corresponding author: Jonathan Gratch, Institute for Creative Technologies, University of Southern California, 12014 Waterfront Drive, Playa Vista, CA 90405, USA.

**Jonathan Gratch** is the Director for Virtual Humans Research at the Institute for Creative Technologies at the University of Southern California (USC) and a research professor in computer science at the USC Viterbi School of Engineering. His e-mail address is [gratch@ict.usc.edu](mailto:gratch@ict.usc.edu).

10.1111/nej.12348

© 2021 President and Fellows of Harvard College

*Negotiation Journal* Winter 2021 13

---

who negotiate their salaries earn considerably more (Marks and Harold 2011; Sève-Söderbergh 2019). It has even been argued that reluctance to negotiate contributes to economic inequality as both women and minorities show greater reluctance to negotiate (Babcock and Laschever 2009; Hernandez and Avery 2016), and even when they negotiate, their demands are met with resistance and backlash (Amanatullah and Tinsley 2013).

What if you had access to an expert negotiator that would make deals on your behalf? Of course, lawyers and personal assistants fulfill this role, but this solution falls outside the means of all but the privileged few. Yet advances in artificial intelligence are raising the possibility of automated personal assistants that can negotiate goods and services on our behalf. Of course, specialized software like eBay bidding bots have been around for a long time, but this technology is rapidly increasing in its sophistication and generality. For example, in the UK, homeowners may soon have their bots negotiate with the bots of power providers to find the best daily electricity prices (Tasquier et al. 2015). More ambitiously, algorithms are emerging that can negotiate directly with customers and sales representatives through natural language. For example, Google famously demonstrated their automated assistant phoning a salon receptionist to negotiate the time of a haircut (Welch 2018), and San Francisco start-up Trim has a chatbot that will negotiate the price of your cable subscription (Farokhmanesh 2016).

In this article, I will discuss the emergence of sophisticated algorithms (henceforth called “agents”) that can directly negotiate with people in human-like ways. This includes agents that use natural communication channels (e.g., text chat or spoken language) as well as those that make use of human psychology to inform, explain, or even manipulate negotiated outcomes. I will highlight recent developments in these technologies and discuss three increasingly ambitious ways in which it will be used. First, automated negotiators are already utilized as research tools that give insight into how people negotiate. Second, they are beginning to be used as automated coaches that help people improve their skills and overcome their reluctance to negotiate. (See Dinnar et al. 2020 in this issue). Finally, automated negotiators have the potential to serve as personal assistants that can negotiate for goods and services on behalf of individuals or companies. This final case, in particular, raises a host of ethical issues that should be considered before this technology reaches widespread use.

## **Automated Negotiators**

The domain of negotiation has been a core focus of artificial intelligence research for the last two decades (Jennings et al. 2001). Much of this

---

research has focused on agents that negotiate with other agents. This work has emphasized precise mathematical formalizations of negotiation (e.g., multi-issue bargaining given some predefined utility function to represent each algorithm's preferences and limits), relied on highly structured mechanisms to manage the negotiation process (e.g., Rubinstein's alternating offers protocol [1982]), based decision making on rational principles (e.g., decision theory and game theory), and to the extent that fairness is even a concern, it is addressed with classical economic principles (e.g., Nash bargaining solution or Shapley values). Most of this research has focused on dyadic negotiation, although there is growing interest in multilateral negotiation (Aydođan et al. 2017) and algorithmic mediation (Rule 2003; Lee and Baykal 2017). Regardless of the number of players, a central goal of this work is the creation of mechanisms to achieve economically efficient (i.e., mutually beneficial) agreements. This is typically cast as a problem of "opponent modeling" (i.e., discovering the other party's priorities and limits) and agents typically exchange thousands of offers per second to explore the range of feasible solutions and find economically efficient and advantageous choices.

Although communication plays a central role in human negotiations, this has been sidestepped by agent-versus-agent approaches to negotiation. Like classical economists, AI researchers often view communication as "cheap talk" that should be ignored. Rather, agents merely exchange sequences of offers (though some research has argued that, even in negotiations between AI agents, efficiency can be improved when they exchange explanations of why offers were rejected; see Rahwan et al. 2003). Communication between agents and the users they represent is also an afterthought; users simply input their preferences and limits through structured menus, though some work has explored more interactive elicitation methods such as conjoint analysis (see Guttman and Maes 1999; Hindriks and Jonker 2008). Users are expected to accept negotiated agreements without explanation as algorithms come with bold guarantees, such as "provably fair" (Goldman and Procaccia 2015).

However, this machine-centric approach to negotiation is now beginning to change. A recent emphasis on "human-aware" artificial intelligence (Kambhampati 2019) has begun to shift the focus of automated negotiation away from rational models and toward a deeper appreciation of behavioral-science research. Human-aware AI seeks to more effectively interact and collaborate with human users by accounting for the unique characteristics of human communication and decision making. This shift has resulted in the emergence of agents that negotiate with people in human-like ways. Research is endowing these agents with the ability to size up their opponents by analyzing their verbal and nonverbal communication (Curhan et al. 2007; Park, Gratch, and

---

Morency 2012; Campos, Alves-Oliveira, and Paiva 2016), to model how people make decisions including common errors and biases (Haim et al. 2010; Stevens et al. 2018; Johnson, Roediger et al. 2019), to communicate through natural language (Rosenfeld et al. 2014, Lewis et al. 2017), to justify their decision making (Crandall et al. 2018, Lee et al. 2019), and to exploit negotiator biases such as fixed-pie assumptions (Nazari, Lucas, and Gratch 2017) or susceptibility to emotional manipulation (de Melo, Carnevale, and Gratch 2011). Before considering how such agents might be used, I briefly discuss these emerging capabilities.

### ***Recognizing Partner State***

Effective negotiators learn about their partner through explicit information exchange (Thompson 1991) but also through observing subtle non-verbal cues (Elfenbein et al. 2007). Artificial negotiators are increasingly acquiring these capabilities. Some approaches focus on the semantic content of a partner's speech. For example, our group has developed techniques to estimate a partner's interests by combining information from their explicit preference statements (e.g., "I prefer a flexible schedule over a high salary") with information gleaned from their pattern of concessions (Nazari, Lucas, and Gratch 2015b). Interestingly, this work found that discrepancies between these channels also predicted deception. In other work, we have used this semantic information to diagnose common errors that novice negotiators make, such as failure to exchange information, susceptibility to anchoring biases, or fixed-pie assumptions (Johnson, Roediger et al. 2019).

Other approaches focus on the manner by which this semantic content is delivered, for example, negotiator facial expressions, gestures, and speech rhythms. Figure One illustrates a multimodal capture and analysis framework that captures and analyzes these signals in real time. Combining these multimodal cues with advanced machine learning methods has allowed us to predict personality traits such as Machiavellianism (Nazari, Lucas, and Gratch 2015a), transitory states such as trustworthiness (Lucas, Stratou, and Gratch 2016), and even whether the partner will accept or reject the most recent offer (Park, Gratch, and Morency 2012).

### ***Modeling Negotiator Decision Making***

Human negotiators do not strictly follow axioms of rationality and are unable to exchange thousands of offers to explore the zone of possible agreements. Rather, they use heuristics, they form mental models of their partner, they give weight to social norms and relational factors like rapport, and they rely on "cheap talk" to guide the process (Bazerman and Carroll 1987). Automated negotiators increasingly model and mimic these uniquely human skills. Algorithms model human-like

**Figure One**

**Illustration of a participant interacting with an automated negotiator (see Gratch et al. 2015). The left side of the image illustrates automatic sensing capabilities including facial expression recognition, posture tracking, and voice analysis (see Stefanov et al. 2020 for the most recent open-source version of this sensing technology). [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]**



theory-of-mind reasoning (Pynadath and Marsella 2005; de Weerd, Broers, and Verbrugge 2015). They use these mental models to discover mutually beneficial trade-offs (Mell, Lucas, and Gratch 2015). They reason about social norms like fairness, and recognize that a person's willingness to accept a deal depends not only on the material outcome, but the perceived fairness of the process and intentions that produced it. For example, Lee and colleagues showed that "provably fair" solutions are rejected by human decision makers unless these solutions match human intuition, yet these intuitions are formed by information that rational algorithms typically fail to provide (Lee et al. 2019). Instead, algorithms must help people understand the space of possible solutions and how their gains compare to those of their partner. In other words, algorithms must incorporate principles of both distributive and procedural justice.

Algorithms also model the common errors or biases that people bring to negotiations, which can be important in explaining or diagnosing why certain agreements occurred. For example, they can simulate how mental models are shaped by emotion (de Melo et al. 2012). They can capture how negotiators often underestimate the opportunity for mutually beneficial solutions or how they fail to attend to information provided by their partner (Johnson and Gratch 2020). They can model how negotiator behavior changes when certain issues are seen as sacred (Dehghani, Carnevale, and Gratch 2014). Algorithms can even capture

---

how people misinterpret nonverbal cues, for example, by underestimating the potential for emotion regulation (Hoegen et al. 2019).

More broadly, a human-aware approach must account for the fact that human decision makers' actions are not based solely on the utilitarian notions that underlie classical negotiation theory. As highlighted by Curhan and Elfenbein (2006), negotiators care about losing face, they value when their concerns are listened to and acknowledged, they want to project a good reputation and punish those with bad reputations, and they assign genuine value to relationships. Moreover, people are willing to reduce their material outcomes to achieve these other goals. For example, prospective employees may be reluctant to negotiate a high salary for fear of damaging their relationship with their future manager (Artz, Goodall, and Oswald 2018). Human-aware techniques are beginning to explicitly consider these nonutilitarian factors. Examples include explicitly modeling subjective factors, like inequity aversion, in the agent's utility function (Katok and Pavlov 2013; Gratch, Nazari, and Johnson 2016); creating reputation mechanisms (Bolton, Greiner, and Ockenfels 2013); or even exploring ways that machines can establish something like rapport (Gratch and Lucas in press).

### ***Modeling Persuasion Tactics***

In that automated negotiators can model how people negotiate and even diagnose common negotiator errors, the possibility arises that algorithms can utilize these models to shape negotiation outcomes. This can be used to benefit human users, for example by helping to convince them to accept a nonintuitive but mutually advantageous deal (Lee et al. 2019) or by creating learning opportunities in a negotiation course (VanLehn et al. 2003; Kapur 2008). However, it can also be used to exploit human weaknesses. As a simple example, automated negotiators can extract greater concessions by making ambitious initial offers (Roediger 2018) or by appearing angry (de Melo, Carnevale, and Gratch 2011). Agents can also use more elaborate strategies, such as building a sense of rapport (Gratch, Kang, and Wang 2014), using favor exchange (Mell, Lucas, and Gratch 2015), or claiming an issue of little importance is quite important, and trading this apparent major concession for something that is actually important (Gratch, Nazari, and Johnson 2016).

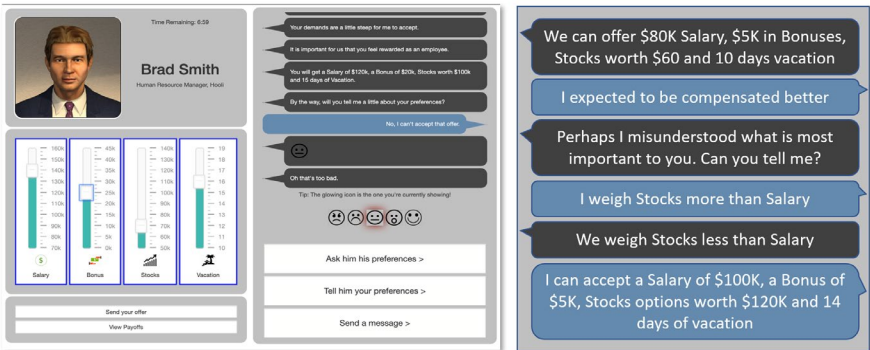
Research shows that these persuasion tactics can be highly effective and allow agents to claim a disproportionate piece of the negotiated pie without seeming unfair. At a recent AI negotiation competition, for instance, prizes went to the team whose agent extracted the greatest concessions from human negotiators. Most contestants treated the competition as a trade-off between exploitation and backlash: as their agents demanded more, they reached a point where human partners would

become angry and reject the negotiation. However, the winning agent found a way to exploit human psychology to transcend this apparent trade-off. Employing what is sometimes called a bogey tactic, the agent pretended to highly value low-value issues, then offered concessions on these issues to get what it truly wanted. As the human negotiators believed the agent was making real concessions, this deceptive strategy left them feeling that the agent was fair and honest, even though its offers were the most unfair of all (Mell et al. 2018).

**Bringing it Together**

No automated negotiator currently brings together the full range of aforementioned capabilities into a single system, but a number of teams are working toward this goal. Figure Two illustrates one of the more widely used systems, our own IAGO negotiation framework. IAGO is a publicly available online platform that allows researchers to define various negotiation scenarios and automated negotiator behaviors, and test these behaviors on human participants (<https://myiago.com/>). Currently, it serves as the underlying software for the Automated Negotiation Agents Competition (ANAC), a yearly competition designed to advance the state of automated negotiators (Mell et al. 2018). Figure Two illustrates a dyadic salary negotiation implemented within the framework.

**Figure Two**  
**Illustration of a dyadic four-issue salary negotiation exercise implemented in the IAGO negotiation agent platform. The left image shows the interface, including menus to exchange offers, messages and preference information, and a history of the dialog. The right image highlights a dialog. The automated hiring manager begins with a low offer. The employee engages in some information exchange before making a counteroffer. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]**



---

IAGO currently sidesteps the problem of natural language communication by having negotiators express their speech acts through menus, but it models many of the cognitive processes and even biases that human negotiators exhibit. Human negotiators can exchange offers with the agent through sliders. They can exchange information about how they rank different issues, asserting a preference for stocks over salary or asking the automated hiring manager about its relative preferences over pairs of issues. Users can also express an emotional attitude toward the negotiation through emojis, or select from a number of canned statements such as “I expect to be compensated better” or “We should consider each other’s interests.”

Different agents are provided with the framework and designers can adapt these or create their own. The default agent aims to play in a human-like way, which means it implements common biases found in human negotiators. For example, although the agent tries to make efficient offers, it struggles to find mutually beneficial solutions. It can find good solutions when its negotiation partner freely exchanges information—in Figure Two, the user’s assertion that they prefer stocks allows the agent to realize it can trade off stocks in exchange for salary concessions. But just like people, this discovery process is shaped by biases. Specifically, the agent begins negotiation with fixed-pie prior (i.e., absent information to the contrary, the agent assumes the participant ranks issues the same way as the agent). Thus, if the user fails to exchange information, the agent will fail to explore trade-offs. Yet other agents successfully exploit human biases, such as tendencies to concede to angry partners (Mell and Gratch 2017), or to partners that anchor the negotiation around a strong initial offer (as in Galinsky and Mussweiler 2001).

The flexibility of the IAGO framework makes it an ideal tool to explore various uses of automated negotiators. We now turn to research that illustrates some obvious cases.

## **Automated Negotiators as Methodological Tools**

Automated tutors have the potential to revolutionize negotiation research. Of course, algorithms for measuring facial expressions or language are playing an increasingly important role in the behavioral sciences (Curhan et al. 2007; Taylor and Thomas 2008). But automated negotiators can create new research opportunities by serving as experimental confederates (i.e., individuals that seem to be participants in a study but are actually enacting theoretically posited behaviors in a research study (see Blascovich et al. 2002; de Melo, Carnevale, and Gratch 2014). Extant research into negotiation processes has adopted one of two experimental methods—the “dyadic approach” and the “scripted



---

approach”—each with its own methodological concerns. I argue that automated negotiators offer a third way that addresses the weaknesses of existing methods.

The first approach, the dyadic approach, allows two (or more) participants to negotiate with each other. For example, in a salary negotiation, one would play the role of a hiring manager and the other play the role of the prospective employee. Studies using such scenarios have found that women negotiate worse salaries when compared with men, particularly when the job is described with gender-threatening language (Kray, Thompson, and Galinsky 2001; Tellhed and Björklund 2011). The advantage of the dyadic approach is it allows for a rich and natural give-and-take between participants (analogous to what they would face in a real salary negotiation) and facilitates study of emergent processes such as information exchange and value creation (Thompson 1991). Its disadvantage is it is difficult to attribute causality for poor outcomes. Do women negotiate lower salaries because they are poor negotiators or because their partners treat them differently than men (Amanatullah and Tinsley 2013)? Such questions might be addressed using human confederates (e.g., having a lab manager play the role of the hiring manager and instruct them to treat men and women equally), but it is difficult to rule out the potential for implicit bias.

To deal with the lack of experimental control inherent in the dyadic approach, many researchers have adopted a “scripted approach” wherein one partner’s behavior is dictated by a negotiation script, though this can introduce other problems that I argue automated negotiations can address. As an example of the scripted approach, in a salary negotiation the hiring manager can be replaced with a completely deterministic computer program that makes the same sequence of statements and concessions, regardless of the participant’s gender or behavior (e.g., Amanatullah and Morris 2010). The advantage of this approach is the hiring manager is literally blind to the participant’s gender and other characteristics, thus any differences can be safely attributed to the employee. However, this increase in experimental control comes at a great cost. One of the hallmarks of strong negotiators is the ability to adapt to one’s partner—to understand their opponent’s interests, communicate their own, and guide the negotiation toward mutually beneficial trade-offs (Thompson 1991; Marks and Harold 2011). Study of such interactive processes is simply not possible if one side is controlled by a deterministic script.

Deterministic scripts can also shape social impressions in unintended ways. For example, most of the research on anger in negotiations relies on scripted agents (e.g., van Kleef, de Dreu, and Manstead 2004). In such studies, the agent follows a fixed concession strategy

---

and, depending on the experimental condition, always expresses either anger or happiness. However, this deterministic use of emotions disrupts the contingencies that give emotions their meaning. In real negotiations, the meaning of an expression is highly contextual—it depends on the events leading up to the expression (de Melo et al. 2014). Thus, anger seems a natural response if the participant has just made an unfair offer, but seems bizarre if the participant has just made a large concession. As a consequence, scripted agents may be effective at controlling behavior, but may fail to control how these behaviors are interpreted.

A broader concern with scripted approaches is that they may fail the “Turing Test” (Turing 2009). In other words, participants become wise to the fact they are merely interacting with a computer. This is important because research shows that people treat computers quite differently from how they treat human partners. For example, people get less angry and are more willing to accept unfair offers from computers (Sanfey et al. 2003) and show less concern for social norms such as fairness (de Melo, Carnevale, and Gratch 2013). Instead, agents that respond in unscripted and intelligent ways are more likely to be treated like people (Blascovich et al. 2002).

In contrast to scripted agents, automated negotiators implement the intelligent and interactive processes found in dyadic negotiations. An example of such interactivity is the ability of agents to form models of their partner’s goals, discover opportunities for trade-offs, and propose efficient solutions based on what information their human partner reveals during the negotiation. This then allows researchers to study how changes to information exchange might alter outcomes. In one study, Roedinger had participants negotiate with an agent that adopted a fixed-pie bias (i.e., the agent began the negotiation with a prior belief that the negotiation was zero-sum). She found that this reinforced participants’ own fixed-pie biases and made it harder for them to find mutually beneficial agreements, as compared to participants who negotiated with an unbiased agent (Roediger 2018).

Using such agents forces a change in mindset around the idea of experimental control. Whereas researchers using scripted agents typically think about controlling specific behaviors (e.g., controlling the sequence of concessions or emotions expressed), researchers using interactive agents must think at the level of interaction mechanisms. So rather than showing a fixed sequence of offers, researchers might study different logrolling strategies. And rather than showing a fixed sequence of emotion, researchers might study the impact of emotional mimicry (Hoegen et al. 2018). Thus, I argue that negotiation agents combine the strengths of both dyadic and scripted approaches. Like the dyadic approach, they support the study of interactive processes, but like the

---

scripted approach, they allow for strong experimental control over the factors that might shape outcomes.

## **Automated Negotiators as Online Tutors**

Automated negotiators have the potential to transform negotiation-skills training. Intelligent tutoring systems already have shown impressive results in teaching hard skills including math (Koedinger et al. 1997; Koedinger and Corbett 2005; Koedinger, Corbett, and others 2006), reading (Wijekumar, Meyer, and Spielvogel 2005; Mills-Tettey et al. 2009), and computer literacy (Guo 2015; Olney et al. 2017). Although still in the early days, research suggests that these methods have promise for teaching interpersonal skills such as negotiation (Kim et al. 2009; Kapp, Blair, and Mesch 2014; Gratch, DeVault, and Lucas 2016; Monahan et al. 2018). Intelligent tutoring systems mirror the way that negotiation is commonly taught. In a negotiation course, the instructor lectures on some principle, students pair up and engage in a negotiation that illustrates that principle, and finally, the instructor provides feedback based on how groups performed. In intelligent tutoring systems, the software “lectures” on some topic (e.g., it might play a video of some algebra principle, like distributing across parentheses), allows students to engage in problems that illustrate the principle (i.e., solve a bunch of math problems), and then automatically provides each student specific feedback based on their performance.

More specifically, tutoring systems work by a principle of model-based diagnosis. For example, rather than simply detecting if a student has the right or wrong answer, these systems explicitly model different classes of common errors that students might make. In algebra, this might include distributing a power, as when a student concludes  $(a + b)^2 = a^2 + b^2$ , or failing to distribute minus signs, as when a student concludes  $-(x + 2y - 5) = -x + 2y - 5$ . Model-based diagnosis takes as input a series of math problems and outputs the model that best explains that student’s pattern of errors. Following this diagnosis, the automated tutor provides feedback personalized for the students’ particular weaknesses (e.g., reviewing the concept of how powers are distributed) and subsequently tests students on problems on which they struggled.

Our work on negotiation tutors adopts the same approach. Following a lecture (for example, on how to exchange information to create mutually beneficial trade-offs), students are allowed to practice negotiating with an automated agent. Model-based diagnosis is then run over the negotiation trace. For example, consider that a student fails to discover a mutually beneficial trade-off. This could be because they failed to ask about their partner’s priorities. It could also be because they failed to make good use of the information that they did receive. Work in our

---

lab has shown that algorithms can automatically diagnose specific errors that novice negotiators make (Johnson, Roediger et al. 2019) and provide feedback tailored to those specific weaknesses. Studies have demonstrated that students improve their performance in subsequent automated negotiations (Monahan et al. 2018; Johnson, Lucas et al. 2019), though research is still needed to see if these benefits generalize to interactions with other people.

More work is needed to realize the full promise of this technology. For example, our tutoring work heavily borrows from the utilitarian perspective that underlies classical negotiation theory (i.e., eliciting the partner's utility function, generating Pareto efficient solutions). It does not account for other sources of value, such as preserving face or establishing a long-term relationship. Nor does it address the communicative aspects of negotiation, such as "reading" or regulating emotions, how to establish rapport, or how to change behavior based on the situational context (e.g., humor may be an effective tactic in some situations but not others). Other learning technology has addressed these communicative skills, though typically outside the context of negotiation. For example, Hoque and colleagues have developed algorithms to make speakers aware of their mannerisms in public speaking and job interviews (e.g., Tanveer et al. 2016), the US Army uses interactive agents to train leaders on how to interview soldiers about suspected sexual harassment (Conant 2018), and Tailspin Reality Labs allows you to practice how to fire an employee and then provides feedback on your performance (Dean 2019). Some companies even offer to monitor and analyze the actual digital communication of work teams and provide automated coaching to team leaders (e.g., see ELIN Inc. or Cultivate.com). As these disparate capabilities begin to coalesce, the potential for online training seems transformative.

## **Automated Negotiators as Proxies**

For those who find negotiation aversive, perhaps the most seductive possibility is for automated negotiators simply to negotiate on their behalf. Many might find it tempting to have an agent negotiate one's salary, and for the 50 percent that fail to negotiate their salaries, this seems like it could only improve their situation. Companies might find it attractive to have agents negotiate the salaries of employees to increase transparency and minimize bias, or simply because they believe it will benefit their bottom line. Indeed, companies routinely screen applicants using artificial intelligence (Cohen 2019) and some companies are even proposing that interviews should be performed and assessed by automated agents (Skantze 2019). Research on "algorithmic bias" (Hajian, Bonchi, and Castillo 2016) suggests the altruistic benefits of this automation

---

should be met with skepticism, but the economic potential will likely drive the deployment of automated negotiators, including their potential to extract greater economic efficiency from transactions, reduce corporate payrolls, and provide convenience for customers.

Delegating negotiations to a machine has potential benefits but raises a host of ethical concerns. For example, our research has found that manipulative negotiation tactics are highly effective (de Melo, Carnevale, and Gratch 2011; Nazari, Lucas, and Gratch 2017). Perhaps more concerning, our research suggests people want their agents to be deceptive. For example, a recent study found that people prefer their automated negotiators to employ ethically questionable tactics such as withholding important information and using explicit deception and emotional manipulation (Mell et al. 2020). This is consistent with a general pattern that suggests people act less ethically when acting through intermediaries, regardless of whether they are human or machine. Research into human organizations highlights the potential for moral concerns to fade to the background when people delegate decisions to human intermediaries. For example, research on *bounded ethicality* suggests that people feel less accountability and show less consideration for others when acting through human agents (Chugh, Bazerman, and Banaji 2005). As a result, people feel comfortable allowing their representatives to make less ethical decisions and adopt riskier courses of action than they themselves might choose (Keltner, Gruenfeld, and Anderson 2003; Rees, Tenbrunsel, and Bazerman 2019). Consistent with this, “technology-mediated” interactions show that people may make riskier decisions or dehumanize others when acting through technology. For example, drone operators exhibit far lower rates of PTSD than combatants in the field (Chappelle et al. 2014), although they are not immune to other forms of stress. The implication is that the act of remote killing is less psychologically distressing than conventional combat. Supporting this, other research reports that medical doctors are more willing to recommend risky and painful procedures (Lee, Fruchter, and Dabbish 2015) and college students are more willing to kill ladybugs (Rutchick et al. 2017) when making these decisions via video rather than in person (though, importantly, individuals differ in their susceptibility to these effects). Taken together, these findings suggest that technological mediation creates a form of psychological distance (Lieberman, Trope, and Stephan 2007) that lowers moral sensitivity to others.

Given that automated negotiators will benefit their clients when they use ethically questionable behavior, and given that clients are psychologically predisposed to allow these tactics, who should decide which tactics these algorithms should adopt? Of course, this challenge is not unique to automated negotiators. Ethicists have struggled recently

---

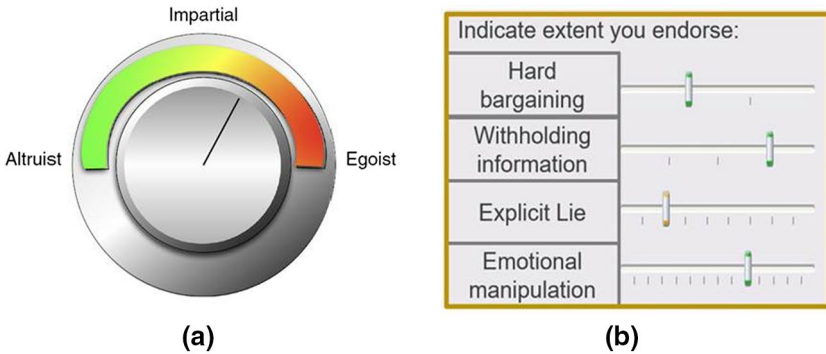
with the dilemmas and legal issues arising from a variety of autonomous systems. For example, cars could be programmed to prioritize either the safety of passengers or the safety of pedestrians. Perhaps not surprisingly, drivers prefer riding in cars that would protect them (Bonneton, Shariff, and Rahwan 2016). This has led some ethicists to conclude that the same ethical principles should be imposed across the board (Gogoll and Müller 2017). Unfortunately, there is little consensus on what constitutes fair or ethical behavior. Even an intuitive concept like fairness can be interpreted in multiple ways: for example, through the *principle of equality* (each gets an equal share), the *principle of equity* (benefits should be distributed in proportion to an individual's investment), or the *principle of need* (benefits should go to those with greatest need). Personality and culture shape which principles a person will tend to adopt (Stouten, de Cremer, and van Dijk 2005), and the same person might evoke different principles in different situations (Wang, Krumhuber, and Gratch 2018). Moreover, the concept of fairness itself can be negotiated—individuals may change their principles in response to arguments or to knowledge about the norms of others (Fehr and Gächter 2000).

To deal with this and other complexities, some researchers have argued that each consumer should have personal control over and responsibility for their agents' behavior. For example, Contissa, Lagioia, and Sartor (2017) have argued that autonomous systems should be fitted with an “ethical knob,” illustrated in Figure Three(a), by which consumers can enact control over their agents. In our own work on negotiation, users were offered the “ethical sliders” illustrated in Figure Three(b), and were quite comfortable sliding them to the right (i.e., encouraging their agent to lie and manipulate).

Having such a knob might seem shocking to some but it simply makes explicit a choice that otherwise may be made without critical thought or examination by the designers or consumers of such systems. For example, one common AI design approach is to automatically learn models by observing human behavior. Facebook used this approach to train automated negotiators. An unintended consequence was the agents learned to lie (Clark 2017). Rather than learn from a population of negotiators, as Facebook did, automated negotiators could learn to copy the consumer's individual style of negotiation. But it is unclear if consumers want agents simply to copy their own behavior, particularly if they are uncomfortable or unskilled at negotiation, or uncomfortable with their own ethics. Indeed, prior work suggests that when people are allowed to reflect critically on their choices, they devise different rules of behavior than they themselves would follow in the moment (Casari and Cason 2009; de Melo, Marsella, and Gratch 2019).

**Figure Three**

**Illustration of the concept of an “ethical knob” that users could use to guide the ethical decision-making of their proxies. The image on the left (a) illustrates the original ethical knob proposed by Contissa and colleagues (2017), whereas the image on the right (b) illustrates a version of this knob that we have adapted for automated negotiators. [Colour figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]**



My own view on this question is evolving but leans toward the “contract model” in legal ethics. As advocated by Peppet (2004), the contract model argues that “lawyers and clients could contractually choose the ethical obligations under which they wanted to operate.” Peppet argues that allowing parties to choose their own ethical constraints will lead to greater efficiencies and greater satisfaction for both negotiators and their clients. When applying this model to autonomy, autonomous negotiators should be required to explicitly disclose, and possibly negotiate, the ethical principles to which they agree to adhere during negotiations.

## Conclusion

Artificial intelligence has made impressive strides in automating low-level skills but we are seeing the arrival of systems with true social intelligence. In this article, I discussed the emergence of agents that can negotiate with people in human-like ways. This includes the ability to diagnose, simulate, and exploit human biases in how people negotiate. These systems have great promise as teachers (helping students practice and improve their skills) and research tools (allowing new forms of experimentation and novel insights into how people negotiate). They also hold promise as tools to increase the convenience and efficiency of economic transactions. Yet, like many technological innovations, they bring new perils. I illustrated how these methods can take advantage of

---

human biases and deploy these advantages at scale. Indeed, research suggests that people are comfortable with their own agents exploiting these weaknesses—when given the opportunity, people instruct their agents to lie and use emotional manipulation to get a better deal. Legal frameworks already govern some of these behaviors, at least for human actors, and could be adapted to sanction unethical behavior by automated negotiators. For example, the American Bar Association has a code of ethics that governs what sorts of deception or manipulation are permissible when negotiating on behalf of clients, and this could be adapted to automated negotiators. But even if such rules make sense, there is no mechanism for enforcing compliance with them. Should ethical constraints be applied by the government; by the companies that create these tools; by professional societies like the IEEE, which guides engineering practices; by consumers? Even without consideration of illegal behavior, there is a wide spectrum of behaviors that some would consider ethically problematic. Sorting out the ethics of this technology will be a challenge for our age.

## Acknowledgments

I am thankful for the thoughtful comments and feedback by Mike Wheeler and the other panelists and participants of the Working Conference on AI, Technology, and Negotiation held by the Harvard Program on Negotiation in May 2020, where these ideas were first presented. This work would not have been possible but for the creative research contributions of my current and former PhD students Celso de Melo, Emmanuel Johnson, and Johnathan Mell, and current and past collaborators Corinne Bendersky, Peter Carnevale, David DeVault, and Peter Kim. Research was sponsored by the Army Research Office under Cooperative Agreement Number W911NF-20-2-0053, the Air Force Office of Scientific Research under grant FA9550-18-1-0182, and the National Science Foundation under grant 1822876. The views and conclusions contained in this document are those of the author 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

- Amanatullah, E. T., and M. W. Morris. 2010. Negotiating gender roles: Gender differences in assertive negotiating are mediated by women's fear of backlash and attenuated when negotiating on behalf of others. *Journal of Personality and Social Psychology* 98(2): 256–267.
- Amanatullah, E. T., and C. H. Tinsley. 2013. Punishing female negotiators for asserting too much...or not enough: Exploring why advocacy moderates backlash against assertive



- female negotiators. *Organizational Behavior and Human Decision Processes* 120(1): 110–122. <https://doi.org/10.1016/j.obhdp.2012.03.006>.
- Artz, B., A. H. Goodall, and A. J. Oswald. 2018. Do women ask? *Industrial Relations: A Journal of Economy and Society* 57(4): 611–636. <https://doi.org/10.1111/irel.12214>.
- Aydođan, R., D. Festen, K. V. Hindriks, and C. M. Jonker. 2017. Alternating offers protocols for multilateral negotiation. In *Modern approaches to agent-based complex automated negotiation*, edited by K. Fujita, Q. Bai, T. Ito, M. Zhang, F. Ren, R. Aydogan, and R. Hadfi, 153–167. New York: Springer.
- Babcock, L., and S. Laschever. 2009. *Women don't ask: Negotiation and the gender divide*. Princeton, NJ: Princeton University Press.
- Bazerman, M. H., and J. S. Carroll. 1987. Negotiator cognition. *Research in Organizational Behavior* 9: 247–288.
- Blascovich, J., J. Loomis, A. Beall, K. Swinth, C. Hoyt, and J. N. Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry* 13(2): 103–124.
- Bolton, G., B. Greiner, and A. Ockenfels. 2013. Engineering trust: Reciprocity in the production of reputation information. *Management Science* 59(2): 265–285. <https://doi.org/10.1287/mnsc.1120.1609>.
- Bonnefon, J.-F., A. Shariff, and I. Rahwan. 2016. The social dilemma of autonomous vehicles. *Science* 352(6293): 1573–1576. <https://doi.org/10.1126/science.aaf2654>.
- Campos, J., P. Alves-Oliveira, and A. Paiva. 2016. Looking for conflict: Gaze dynamics in a dyadic mixed-motive game. *Autonomous Agents and Multi-Agent Systems* 30(1): 112–135.
- Casari, M., and T. N. Cason. 2009. The strategy method lowers measured trustworthy behavior. *Economics Letters* 103(3): 157–159. <https://doi.org/10.1016/j.econlet.2009.03.012>.
- Chappelle, W. L., K. D. McDonald, L. Prince, T. Goodman, B. N. Ray-Sannerud, and W. Thompson. 2014. Symptoms of psychological distress and post-traumatic stress disorder in United States Air Force “drone” operators. *Military Medicine* 179(suppl\_8): 63–70. <https://doi.org/10.7205/milmed-d-13-00501>.
- Chugh, D., M. H. Bazerman, and M. R. Banaji. 2005. Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy*, edited by D. A. Moore, D. M. Cain, G. Loewenstein, and M. H. Bazerman, 74–95. New York: Cambridge University Press.
- Clark, L. 2017. Facebook teaches bots how to negotiate. They learn to lie instead. *Wired*, June 15.
- Cohen, T. 2019. How to leverage artificial intelligence to meet your diversity goals. *Strategic HR Review* 18(2): 62–65.
- Conant, J. 2018. Army launches new SHARP training tools. *Defense Visual Information Distribution Service*, March 5.
- Consumer Reports. 2013. Consumer Reports’ survey: Nine of 10 Americans who haggled saved money. Consumer Reports. Available from <https://www.consumerreports.org/media-room/press-releases/2013/07/my-entry/>.
- Contissa, G., F. Lagioia, and G. Sartor. 2017. The ethical knob: Ethically-customisable automated vehicles and the law. *Artificial Intelligence and Law* 25(3): 365–378. <https://doi.org/10.1007/s10506-017-9211-z>.
- Crandall, J. W., M. O. Tennom, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon, M. Cebrian, A. Shariff, M. A. Goodrich, and I. Rahwan. 2018. Cooperating with machines. *Nature Communications* 9(1): 233. <https://doi.org/10.1038/s41467-017-02597-8>.
- Curhan, J. R., and H. A. Elfenbein. 2006. What do people value when they negotiate? Mapping the domain of subjective value in negotiation. *Journal of Personality & Social Psychology* 91(3): 493–512.
- Curhan, J. R., A. Pentland, R. Caneel, N. Eagle, and M. C. Martin. 2007. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first five minutes. *Journal of Applied Psychology* 92(3): 802–811.
- de Melo, C., P. J. Carnevale, and J. Gratch. 2011. The effect of expression of anger and happiness in computer agents on negotiations with humans. Presented at the 10th International Conference on Autonomous Agents and Multiagent Systems, Taipai, Taiwan.
- \_\_\_\_\_. 2013. Agent or avatar? Using virtual confederates in conflict management research. Presented at the 73rd Annual Meeting of the Academy of Management, Orlando, FL.

- \_\_\_\_\_. 2014. Using virtual confederates to research intergroup bias and conflict. Presented at the 74th Annual Meeting of the Academy of Management, Philadelphia, PA.
- de Melo, C., P. J. Carnevale, S. J. Read, and J. Gratch. 2014. Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology* 106(1): 73–88. <https://doi.org/10.1037/a0034251>.
- de Melo, C. M., P. Carnevale, S. Read, D. Antos, and J. Gratch. 2012. Bayesian model of the social effects of emotion in decision-making in multiagent systems. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems* (Vol. 1). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
- de Melo, C., S. Marsella, and J. Gratch. 2019. Human cooperation when acting through autonomous machines. In *Proceedings of the National Academy of Sciences* 116(9): 3482–3487. <https://doi.org/10.1073/pnas.1817656116>.
- de Weerd, H., E. Broers, and R. Verbrugge. 2015. Savvy software agents can encourage the use of second-order theory of mind by negotiators. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*, 542–547. Pasadena, CA: Cognitive Science Society.
- Dean, S. 2019. Barry sobbed as he begged me not to fire him. I canned him anyway, over and over. *Los Angeles Times*, August 27. Available from <https://www.latimes.com/business/story/2019-08-26/virtual-reality-vr-technology-fire-simulator-future>.
- Dehghani, M., P. J. Carnevale, and J. Gratch. 2014. Interpersonal effects of expressed anger and sorrow in morally charged negotiation. *Judgment and Decision Making* 9(2): 104–113.
- Dinnar, S. M., C. Dede, E. Johnson, C. Straub, and K. Korjus. 2020. Artificial intelligence and technology in teaching negotiation. *Negotiation Journal* 37(1): 65–82.
- Elfenbein, H. A., M. Der Foo, J. White, H. H. Tan, and V. C. Aik. 2007. Reading your counterpart: The benefit of emotion recognition accuracy for effectiveness in negotiation. *Journal of Nonverbal Behavior* 31(4): 205–223. <https://doi.org/10.1007/s10919-007-0033-7>.
- Farokhmanesh, M. 2016. Stop arguing with Comcast and let this bot negotiate for you. *The Verge*, November 17.
- Fehr, E., and S. Gächter. 2000. Fairness and retaliation: The economics of reciprocity. *Journal of Economic Perspectives* 14(3): 159–181. <https://doi.org/10.1257/jep.14.3.159>.
- Galinsky, A. D., and T. Mussweiler. 2001. First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology* 81(4): 657–669. <https://doi.org/10.1037/0022-3514.81.4.657>.
- Glassdoor Team. 2016. 3 in 5 employees did not negotiate salary [Blog Post]. In *the News*, May 2. Available from <https://www.glassdoor.com/blog/3-5-u-s-employees-negotiate-salary/>.
- Gogoll, J., and J. F. Müller. 2017. Autonomous cars: In favor of a mandatory ethics setting. *Science and Engineering Ethics* 23(3): 681–700. <https://doi.org/10.1007/s11948-016-9806-x>.
- Goldman, J., and A. D. Proccaccia. 2015. Spliddit: Unleashing fair division algorithms. *SIGecom Exchanges* 13(2): 41–46. <https://doi.org/10.1145/2728732.2728738>.
- Gratch, J., D. DeVault, and G. Lucas. 2016. The benefits of virtual humans for teaching negotiation. In *Proceedings of the 16th International Conference on Intelligent Virtual Agents (IVA), 2016, Los Angeles, CA, USA*. New York: Springer.
- Gratch, J., D. DeVault, G. Lucas, and S. Marsella. 2015. Negotiation as a challenge problem for virtual humans. Presented at the 15th International Conference on Intelligent Virtual Agents, Delft, the Netherlands.
- Gratch, J., S.-H. Kang, and N. Wang. 2014. Using social agents to explore theories of rapport and emotional resonance. In *Social emotions in nature and artifact*, edited by J. Gratch and S. Marsella. New York: Oxford University Press.
- Gratch, J., and G. M. Lucas. in press. Rapport between humans and socially interactive agents. In *Handbook of socially interactive agents*, edited by B. Lugrin, C. Pelachaud, and D. Traum. San Rafael, CA: ACM Books.
- Gratch, J., Z. Nazari, and E. Johnson. 2016. The misrepresentation game: How to win at negotiation while seeming like a nice guy. Presented at the International Conference on Autonomous Agents and Multiagent Systems, Singapore.
- Guo, P. J. 2015. Codeopticon: Real-time, one-to-many human tutoring for computer programming. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology*. New York: Assoc. for Computing Machinery.
- Guttman, R. H., and P. Maes. 1999. Agent-mediated integrative negotiation for retail electronic commerce. In *AMET '98: Selected Papers from the First International Workshop*

- 
- on Agent Mediated Electronic Trading on Agent Mediated Electronic Commerce, Berlin, Heidelberg, 70–90. New York: Assoc. for Computing Machinery.
- Haim, G., Y. Gal, S. Kraus, and Y. Blumberg. 2010. Learning human negotiation behavior across cultures. Presented at the HuCom10-Second International Working Conference on Human Factors and Computational Models in Negotiation, Delft, the Netherlands.
- Hajian, S., F. Bonchi, and C. Castillo. 2016. Algorithmic bias: From discrimination discovery to fairness-aware data mining. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA*, 2125–2126. New York: Assoc. for Computing Machinery.
- Hernandez, M., and D. R. Avery. 2016. Getting the short end of the stick: Racial bias in salary negotiations. *MIT Sloan Management Review*, June 15. Available from <https://sloanreview.mit.edu/article/getting-the-short-end-of-the-stick-racial-bias-in-salary-negotiations/>.
- Hindriks, K. V., and C. M. Jonker. 2008. Creating human-machine synergy in negotiation support systems: Towards the pocket negotiator. In *Proceedings of the 1st International Working Conference on Human Factors and Computational Models in Negotiation*. New York: Assoc. for Computing Machinery.
- Hoegen, R., J. Gratch, B. Parkinson, and D. Shore. 2019. Signals of emotion regulation in a social dilemma: Detection from face and context. Presented at the 8th International Conference on Affective Computing & Intelligent Interaction, Cambridge, UK.
- Hoegen, R., J. van der Schalk, G. Lucas, and J. Gratch. 2018. The impact of agent facial mimicry on social behavior in a prisoner's dilemma. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents, Sydney, NSW, Australia*. New York: Assoc. for Computing Machinery.
- Jennings, N. R., P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. Wooldridge. 2001. Automated negotiation: Prospects, methods and challenges. *International Journal of Group Decision and Negotiation* 10(2): 199–215.
- Johnson, E., and J. Gratch. 2020. The impact of implicit information exchange in human-agent negotiations. Presented at the International Conference on Intelligent Virtual Agents, Glasgow, Scotland.
- Johnson, E., G. Lucas, P. Kim, and J. Gratch. 2019. Intelligent tutoring system for negotiation skills training. Presented at the 20th International Conference on Artificial Intelligence in Education, Chicago, IL.
- Johnson, E., S. Roediger, G. Lucas, and J. Gratch. 2019. Assessing common errors students make when negotiating. *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents, Paris, France*. New York: Assoc. for Computing Machinery.
- Kambhampati, S. 2019. Challenges of human-aware AI systems. *arXiv preprint*. arXiv:1910.07089.
- Kapp, K. M., L. Blair, and R. Mesch. 2014. *The gamification of learning and instruction field-book: Ideas into practice*. San Francisco, CA: Wiley.
- Kapur, M. 2008. Productive failure. *Cognition and Instruction* 26(3): 379–424. <https://doi.org/10.1080/07370000802212669>.
- Katok, E., and V. Pavlov. 2013. Fairness in supply chain contracts: A laboratory study. *Journal of Operations Management* 31(3): 129–137.
- Keltner, D., D. H. Gruenfeld, and C. Anderson. 2003. Power, approach, and inhibition. *Psychological Review* 110(2): 265–284. <https://doi.org/10.1037/0033-295X.110.2.265>.
- Kim, J. M., R. W. Hill, P. J. Durlach, H. C. Lane, E. Forbell, M. Core, S. Marsella, D. Pynadath, and J. Hart. 2009. BILAT: A game-based environment for practicing negotiation in a cultural context. *International Journal of Artificial Intelligence in Education* 19: 289–308.
- Koedinger, K., and A. Corbett. 2005. Cognitive tutors. In *The Cambridge handbook of the learning sciences*, edited by R. K. Sawyer, 61–78. New York: Cambridge University Press.
- Koedinger, K. R., J. R. Anderson, W. H. Hadley, and M. A. Mark. 1997. Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education (IJAIED)* 8: 30–43.
- Koedinger, K. R., A. Corbett, and others. 2006. *Cognitive tutors: Technology bringing learning sciences to the classroom*. Available from [https://www.researchgate.net/publication/n/255519934\\_Cognitive\\_Tutors\\_Technology\\_Bringing\\_Learning\\_Science\\_to\\_the\\_Classroom](https://www.researchgate.net/publication/n/255519934_Cognitive_Tutors_Technology_Bringing_Learning_Science_to_the_Classroom).
- Kray, L. J., L. Thompson, and A. Galinsky. 2001. Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology* 80(6): 942–958. <https://doi.org/10.1037/0022-3514.80.6.942>.

- 
- Lee, M. K., and S. Baykal. 2017. Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. New York: Assoc. for Computing Machinery.
- Lee, M. K., N. Fruchter, and L. Dabbish. 2015. Making decisions from a distance: The impact of technological mediation on riskiness and dehumanization. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, Vancouver, BC, Canada*. New York: Assoc. for Computing Machinery.
- Lee, M. K., A. Jain, H. J. Cha, S. Ojha, and D. Kusbit. 2019. Procedural justice in algorithmic fairness: Leveraging transparency and outcome control for fair algorithmic mediation. In *Proceedings of the ACM on Human-Computer Interaction 3 (CSCW)*, 1–26.
- Lewis, M., D. Yarats, Y. N. Dauphin, D. Parikh, and D. Batra. 2017. Deal or no deal? End-to-end learning for negotiation dialogues. *arXiv preprint*. arXiv:1706.05125.
- Liberman, N., Y. Trope, and E. Stephan. 2007. Psychological distance. In *Social psychology: Handbook of basic principles*, 2nd ed., edited by A. W. Kruglanski and E. T. Higgins, 353–381. New York: The Guilford Press.
- Lucas, G., G. Stratou, and J. Gratch. 2016. Trust me: Multimodal signals of trustworthiness for virtual human negotiators. Presented at the International Conference on Multimodal Interaction, Tokyo.
- Marks, M., and C. Harold. 2011. Who asks and who receives in salary negotiation. *Journal of Organizational Behavior* 32(3): 371–394. <https://doi.org/10.1002/job.671>.
- Mell, J., and J. Gratch. 2017. Grumpy and Pinocchio: The effect of language and strategy in human-agent negotiation. Presented at the 16th International Conference on Autonomous Agents and Multiagent Systems, Sao Paulo, Brazil.
- Mell, J., J. Gratch, T. Baarslag, R. Aydrogan, and C. Jonker. 2018. Results of the first annual human-agent league of the automated negotiating agents competition. Presented at the 18th International Conference on Intelligent Virtual Agents, Sydney, Australia.
- Mell, J., G. Lucas, and J. Gratch. 2015. An effective conversation tactic for creating value over repeated negotiations. Presented at the 14th International Conference on Autonomous Agents and Multiagent Systems, Istanbul, Turkey.
- Mell, J., G. Lucas, S. Mozgai, and J. Gratch. 2020. The effects of experience on deception in human-agent negotiation. *Journal of Artificial Intelligence Research* 68: 633–660.
- Mills-Tettey, G. A., J. Mostow, M. B. Dias, T. M. Sweet, S. M. Belousov, M. F. Dias, and H. Gong. 2009. Improving child literacy in Africa: Experiments with an automated reading tutor. Presented at the 2009 International Conference on Information and Communication Technologies and Development (ICTD).
- Monahan, S., E. Johnson, G. Lucas, J. Finch, and J. Gratch. 2018. Autonomous agent that provides automated feedback improves negotiation skills. Presented at the 19th International Conference on Artificial Intelligence in Education, London, UK.
- Nazari, Z., G. Lucas, and J. Gratch. 2015a. Multimodal approach for automatic recognition of Machiavellianism. Presented at the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'an, China.
- \_\_\_\_\_. 2015b. Opponent modeling for virtual human negotiators. Presented at the 15th International Conference on Intelligent Virtual Agents, Delft, the Netherlands.
- \_\_\_\_\_. 2017. Fixed-pie lie in action. Presented at the 17th International Conference on Intelligent Virtual Agents, Stockholm, Sweden.
- Olney, A. M., D. Bakhtiari, D. Greenberg, and A. Graesser. 2017. Assessing computer literacy of adults with low literacy skills. Presented at the International Conference on Educational Data Mining (EDM), Wuhan, China.
- Park, S., J. Gratch, and L. P. Morency. 2012. I already know your answer: Using nonverbal behaviors to predict immediate outcomes in a dyadic negotiation. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction*, 19–22. New York: Assoc. for Computing Machinery.
- Peppet, S. R. 2004. Lawyers' bargaining ethics, contract, and collaboration: The end of the legal profession and the beginning of professional pluralism. *Iowa Law Review* 90: 475–538.
- Pynadath, D. V., and S. C. Marsella. 2005. PsychSim: Modeling theory of mind with decision-theoretic agents. Presented at the 19th International Joint Conference on Artificial Intelligence (IJCAI).

- 
- Rahwan, I., S. D. Ramchurn, N. R. Jennings, P. Mccburney, S. Parsons, and L. Sonenberg. 2003. Argumentation-based negotiation. *Knowledge Engineering Review* 18(4): 343–375.
- Rees, M. R., A. E. Tenbrunsel, and M. H. Bazerman. 2019. Bounded ethicality and ethical fading in negotiations: Understanding unintended unethical behavior. *Academy of Management Perspectives* 33(1): 26–42.
- Roediger, S. 2018. The effect of suspicion on emotional influence tactics in virtual human negotiation. Available from [https://essay.utwente.nl/76800/1/Roediger\\_MA\\_EEMCS.pdf](https://essay.utwente.nl/76800/1/Roediger_MA_EEMCS.pdf).
- Rosenfeld, A., I. Zuckerman, E. Segal-Halevi, O. Drein, and S. Kraus. 2014. NegoChat: A chat-based negotiation agent. Presented at the 13th International Conference on Autonomous Agents and Multi-agent Systems, Paris, France.
- Rubinstein, A. 1982. Perfect equilibrium in a bargaining model. *Econometrica: Journal of the Econometric Society* 50(1): 97–109.
- Rule, C. 2003. *Online dispute resolution for business: B2B, e-commerce, consumer, employment, insurance, and other commercial conflicts*. San Francisco, CA: Jossey-Bass.
- Rutchick, A. M., R. M. McManus, D. M. Barth, R. J. Youmans, A. T. Ainsworth, and H. J. Goukassian. 2017. Technologically facilitated remoteness increases killing behavior. *Journal of Experimental Social Psychology* 73: 147–150. <https://doi.org/10.1016/j.jesp.2017.07.001>.
- Sanfey, A. G., J. K. Rilling, J. A. Aronson, L. E. Nystrom, and J. D. Cohen. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300(5626): 1755–1758. <https://doi.org/10.1126/science.1082976>.
- Säve-Söderbergh, J. 2019. Gender gaps in salary negotiations: Salary requests and starting salaries in the field. *Journal of Economic Behavior & Organization* 161: 35–51. <https://doi.org/10.1016/j.jebo.2019.01.019>.
- Skantze, G. 2019. Can we reduce recruitment bias with social robots? *Furhat blog*, January 8. Available from <https://furhatrobotics.com/blog/can-we-reduce-recruitment-bias-with-social-robots/>.
- Stefanov, K., B. Huang, Z. Li, and M. Soleymani. 2020. OpenSense: A platform for multimodal behavior acquisition and perception. Presented at the 22nd ACM International Conference on Multimodal Interaction.
- Stevens, C. A., J. Daamen, E. Gaudrain, T. Renkema, J. D. Top, F. Cnossen, and N. A. Taatgen. 2018. Using cognitive agents to train negotiation skills. *Frontiers in Psychology* 9(154). <https://doi.org/10.3389/fpsyg.2018.00154>.
- Stouten, J., D. de Cremer, and E. van Dijk. 2005. All is well that ends well, at least for pro-felves: Emotional reactions to equality violation as a function of social value orientation. *European Journal of Social Psychology* 35(6): 767–783. <https://doi.org/10.1002/ejsp.276>.
- Tanveer, M. I., R. Zhao, K. Chen, Z. Tiet, and M. E. Hoque. 2016. Automanner: An automated interface for making public speakers aware of their mannerisms, In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 385–396. New York: Assoc. for Computing Machinery.
- Tasquier, L., M. Scialdone, R. Aversa, and S. Venticinque. 2015. Agent based negotiation of decentralized energy production. *Studies in Computational Intelligence* 570: 59–67.
- Taylor, P. J., and S. Thomas. 2008. Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research* 1(3): 263–281. <https://doi.org/10.1111/j.1750-4716.2008.00016.x>.
- Tellhed, U., and F. Björklund. 2011. Stereotype threat in salary negotiations is mediated by reservation salary. *Scandinavian Journal of Psychology* 52(2): 185–195. <https://doi.org/10.1111/j.1467-9450.2010.00855.x>.
- Thompson, L. L. 1991. Information exchange in negotiation. *Journal of Experimental Social Psychology* 27(2): 161–179. [https://doi.org/10.1016/0022-1031\(91\)90020-7](https://doi.org/10.1016/0022-1031(91)90020-7).
- Turing, A. M. 2009. Computing machinery and intelligence. In *Parsing the Turing test: Philosophical and methodological issues in the quest for the thinking computer*, edited by R. Epstein, G. Roberts, and G. Beber, 23–65. Dordrecht, the Netherlands: Springer.
- van Kleef, G. A., C. de Dreu, and A. S. R. Manstead. 2004. The interpersonal effects of anger and happiness in negotiations. *Journal of Personality and Social Psychology* 86(1): 57–76.
- VanLehn, K., S. Siler, C. Murray, T. Yamauchi, and W. B. Baggett. 2003. Why do only some events cause learning during human tutoring? *Cognition and Instruction* 21(3): 209–249.

- 
- Wang, X., E. G. Krumhuber, and J. Gratch. 2018. The interpersonal effects of emotions in money versus candy games. *Journal of Experimental Social Psychology* 79: 315–327. <https://doi.org/10.1016/j.jesp.2018.08.014>.
- Welch, C. 2018. Google just gave a stunning demo of Assistant making an actual phone call. *The Verge*, May 8. Available from <https://www.theverge.com/2018/5/8/17332070/google-assistant-makes-phone-call-demo-duplex-io-2018>.
- Wijekumar, K., B. Meyer, and J. Spielvogel. 2005. Web-based intelligent tutoring to improve reading comprehension in elementary and middle schools: Design, research, and preliminary findings. Presented at E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education.