“Do As I Say, Not As I Do:” Challenges in Delegating Decisions to Automated Agents

Celso M. de Melo
Institute for Creative Technologies, University of Southern California, 12015 Waterfront Drive, Building #4, Playa Vista, CA 90094-2536
demelo@usc.edu

Stacy Marsella
College of Computer and Information Science, Northeastern University, 440 Huntington Avenue, Boston, Massachusetts 02115
stacymarsella@gmail.com

Jonathan Gratch
Institute for Creative Technologies, University of Southern California, 12015 Waterfront Drive, Building #4, Playa Vista, CA 90094-2536
gratch@ict.usc.edu

ABSTRACT
There has been growing interest, across various domains, in computer agents that can decide on behalf of humans. These agents have the potential to save considerable time and help humans reach better decisions. One implicit assumption, however, is that, as long as the algorithms that simulate decision-making are correct and capture how humans make decisions, humans will treat these agents similarly to other humans. Here we show that interaction with agents that act on our behalf or on behalf of others is richer and more interesting than initially expected. Our results show that, on the one hand, people are more selfish with agents acting on behalf of others, than when interacting directly with others. We propose that agents increase the social distance with others which, subsequently, leads to increased demand. On the other hand, when people task an agent to interact with others, people show more concern for fairness than when interacting directly with others. In this case, higher psychological distance leads people to consider their social image and the long-term consequences of their actions and, thus, behave more fairly. To support these findings, we present an experiment where people engaged in the ultimatum game, either directly or via an agent, with others or agents representing others. We show that these patterns of behavior also occur in a variant of the ultimatum game – the impunity game – where others have minimal power over the final outcome. Finally, we study how social value orientation – i.e., people’s propensity for cooperation – impact these effects. These results have important implications for our understanding of the psychological mechanisms underlying interaction with agents, as well as practical implications for the design of successful agents that act on our behalf or on behalf of others.

Categories and Subject Descriptors
I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence – *Intelligent Agents*; J.4 [**Social and Behavioral Sciences**]: Economics, Psychology

Appears in: *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016), J. Thangarajah, K. Tuyls, C. Jonker, S. Marsella (eds.),*
May 9–13, 2016, Singapore.
Copyright © 2016, International Foundation for Autonomous Agents And Multiagent Systems (www.ifaamas.org). All rights reserved.

General Terms
Experimentation, Economics, Human Factors, Theory.

Keywords
Automated Agents, Decision Making, Social Distance, Social Image, Power, Social Value Orientation

1. INTRODUCTION
Recent times have seen increased interest in automated computational systems that make decisions on behalf of people [1]-[8]. These systems have the potential to save considerable time and effort by helping humans reach optimal decisions in complex negotiations or other economic settings [1]-[5]. They can help business leaders reduce labor cost, improve quality, enforce company policy, and work around the clock [6]. They also have the potential to free us from mundane activities, such as driving [7] and they can deliver merchandise for us very efficiently [8]. For these and other reasons, there has been considerable investment in applying these automated agents in health, finance, retail, driving, manufacturing, robotics, defense, etc.

One implicit assumption when creating these systems is that if these agents implement correct decision algorithms and capture how humans make decisions in similar settings, then people will behave with them similarly to how they do with humans. In fact, initial theories of human-computer interaction suggest that to the extent that agents display social cues (e.g., interactivity, verbal and non-verbal behavior), people will treat them in a fundamentally social manner and apply the same rules they use when interacting with humans [9]-[14]. However, recent studies suggest that people still make important distinctions in their behavior with computers, when compared to humans [15]-[17]. These results show that in the same economic tasks, for the same financial incentives, people make more favorable decisions with humans than computers [18]-[21], and show patterns of higher brain activation with humans [22]-[24]. These studies, though, focus on computers that acted *on their own behalf*; in contrast, here we are interested in people’s behavior with agents that act *on behalf of humans*. The answer to this research question is, comparatively, much less well understood.

This paper, therefore, explicitly tests whether agents that act on behalf of humans are treated in the same manner as humans. To accomplish this, we present an experiment where participants engaged in the ultimatum bargaining and the impunity games, as proposers, with other humans or agents representing other humans. Moreover, participants either made decisions directly with their counterparts or via an automated agent. Our results show that there are fundamental differences in the way people behave with these automated agents, when compared to humans; moreover, people act differently with agents that represent them, when compared to agents that represent others. Specifically, the results show that people will task their agents to behave fairly ("Do as I say"), but will act selfishly with agents that represent others ("Not as I do"). Finally, we look at how power – i.e., the counterpart's ability to shape the final outcome – and social value orientation – i.e., an individual trait that measures people's propensity for cooperation – influence these effects.

1.1 Social Distance & Agents That Act on Behalf of Others

As mentioned above, initial theories of human-computer interaction predict that, in social settings, people will interact with agents in the same manner as they would with humans [5], [9]-[14]. The argument is that in such settings people use cognitive heuristics and apply the rules of human-human interaction in human-agent interaction. Thus, in our case, the prediction is that: People's decisions would be as favorable with agents that act on behalf of humans, as with humans.

Research on the effects of social distance on decision making [25]-[31], however, suggest that people will distinguish between interacting with humans directly vs. interacting with an automated agent that acts on behalf of humans. In this research, social distance between participants and their counterparts was achieved by manipulating anonymity, familiarity with others, temporal distance, and physical distance. Hoffman, McCabe, and Smith [25], [26] showed that in conditions of full anonymity – i.e., participants could not be identified by their counterparts or the experimenters – people offered much less than when anonymity was not preserved. Researchers also showed that people offered more in a dictator game to friends than strangers [27], [28]. Pronin, Olivola, and Kennedy [29] demonstrated that temporal distance also affected decision making; in their studies, people showed larger psychological distance between their future selves, who were treated the same as different people, than their present selves. Finally, various researchers have shown that physical proximity can lead to increased cooperation (e.g., [30], [31]). These results, thus, suggest that people will perceive agents that act on behalf of other humans as being more socially distant than the humans themselves. Therefore, in contrast to the prediction in the previous paragraph, we advanced the following hypothesis:

Hypothesis 1: *People's decisions will be more favorable to humans than to agents that act on behalf of other humans.*

1.2 Social Image & Agents that Act on Our Behalf

When people interact via an agent, the physical distance to their counterparts increases. Thus, a straightforward application of the argument presented in the previous section should lead to the following hypothesis:

Hypothesis 2a: *People's decisions when engaging with others via an agent should be less favorable than when interacting with others directly.*

However, interacting with an agent that represents us is different than interacting with agents that represent others. In contrast to the latter case, when people task an agent to be their representative, they are likely to take special considerations for one's reputation or social image. An automated agent is a persistent representation of the person's values and is likely to engage in multiple interactions with various counterparts. Thus, rather than take solely into account the short-term impact, people will also consider the long-term effect of their choices. For all these reasons, people have added incentive to behave fairly when interacting via an automated agent.

Supporting this view, research in the behavioral sciences shows that people tend to behave more fairly when their actions impact their reputations [32], [33] or are being watched by an audience [34]. In a related line of research, construal level theory [35] argues that people focus on concrete and specific aspects of events or objects that are perceived to be psychologically close; in contrast, people focus on more abstract and global aspects when something is perceived to be more psychologically distant. Accordingly, studies demonstrate that when people adopt an abstract level of construal (the "big picture"), rather than focus on the particular interaction they are involved in – and, thus, seek a short-term selfish reward – they, instead, cooperate more in social and moral dilemmas [36]-[38], as well as in negotiation [39], [40].

Thus, we advance a second competing hypothesis:

Hypothesis 2b: *People's decisions when engaging with others via an agent should be more favorable than when interacting with others directly.*

We next consider how two factors that usually have important consequences in social interaction – power and social value orientation – impact decisions with automated agents.

1.3 Power

Power can be broadly defined as the ability to exert influence on other people [41], [42]. In social decision making, one party often has more power than the other and this leads to important consequences to the way people behave and to the final outcome. Research shows that powerful parties tend to have higher aspirations [43], demand more and concede less [44], and be more likely to use threats to get their way [45]. Because power plays such a critical role, we also wanted to explore how this situational factor moderates people's offers with automated agents. To accomplish this we compared behavior in the ultimatum game – where the responder has considerable power, in that if s/he rejects the offer, no one gets anything – with the impunity game – where the responder has little power, in that s/he can only symbolically reject the offer, without affecting the outcome for the proposer. Since it was not clear, however, how power would interact with the type of automated agents involved, we advanced a research question:

Research Question 1: *How will counterpart power influence people's decisions when they are interacting via or with agents?*

1.4 Social Value Orientation

Whereas power is a situational factor, social value orientation (SVO) is an individual factor that captures one's propensity for cooperation [46], [47]. This trait distinguishes between two types of individuals: cooperators (or prosocials) and non-cooperators (or proselfs). The former tend to cooperate by default, though they are able to adjust their behavior in the presence of non-cooperators. Non-cooperators, on the other hand, tend to act selfishly by default, and often need external incentive to cooperate. SVO is argued to be a stable personality trait that is shaped by biology and one's early history of social interaction. Several studies have demonstrated that this trait has a powerful influence on people's decision making in social settings [48], [49]. In our work, we wanted to study whether this individual trait could impact behavior with automated agents. Thus, we advanced a final research question:

> ***Research Question 2:*** *How will social value orientation influence people's decisions when they are interacting via or with agents?*

2. EXPERIMENT

To test our hypotheses and address our research questions, we ran an experiment where participants engaged in the ultimatum game with automated agents that either represented the participants themselves or the counterparts. To manipulate power, participants also engaged in a variant of the ultimatum game – the impunity game – where counterparts are given little power over the final outcome.

2.1 Methods

Design. The experiment followed a 2 × 2 × 2 mixed factorial design: *Responder* (Human vs. Agent; between-participants) × *Proposer* (Human vs. Agent; between-participants) × *Power* (Ultimatum game vs. Impunity game; within-participants). Participants always assumed the role of proposers.

Tasks. In the ultimatum game [50], there are two players: a proposer and a responder. The proposer is given an initial endowment of money and has to decide how much to offer to the responder. Then, the responder has to make a decision: if the offer is accepted, both players get the proposed allocation; if the offer is rejected, however, no one gets anything. The standard rational prediction is that the proposer should offer the minimum non-zero amount, as the responder will always prefer to have something to nothing. In practice, people usually offer 40 to 50 percent of the initial endowment and low offers (about 20 percent of the endowment) are usually rejected [51]. This behavior is usually explained by a concern with fairness and a fear of being rejected [52].

The impunity game is similar to the ultimatum game [53], [54]. The proposer is given an initial endowment of money and makes an offer to a responder, who must decide whether to accept or reject the offer. The critical difference is that, if the offer is rejected, the responder gets zero, but the proposer still keeps the money s/he designated for her-/himself. A rejection by the responder, thus, does not impact the proposer's payoff and is only symbolic. The impunity game can therefore be seen as a version of the ultimatum game where responders are given less power over the outcome[1]. Experimental results with this game show that proposers tend to offer less than in the ultimatum game, though still above the rational prediction of zero [53].

In our experiment, participants were always assigned the role of proposers. In each game, participants were given an initial endowment of 20 lottery tickets. They could make an offer ranging from 0 to 20 tickets. These tickets had financial consequences as they would enter lotteries (one per game) worth $30. Participants were assigned to one of the four possible Proposer × Responder conditions, and played one round of the ultimatum game and one round of the impunity game. The order with which these games was presented was counterbalanced across participants. Before engaging in the actual games, participants read the instructions, were quizzed on the instructions, and completed a tutorial. The interface was also different for these games in terms of colors and icons on screen to make sure people did not confuse the two games. A snapshot for the ultimatum game is shown in Figure 1.

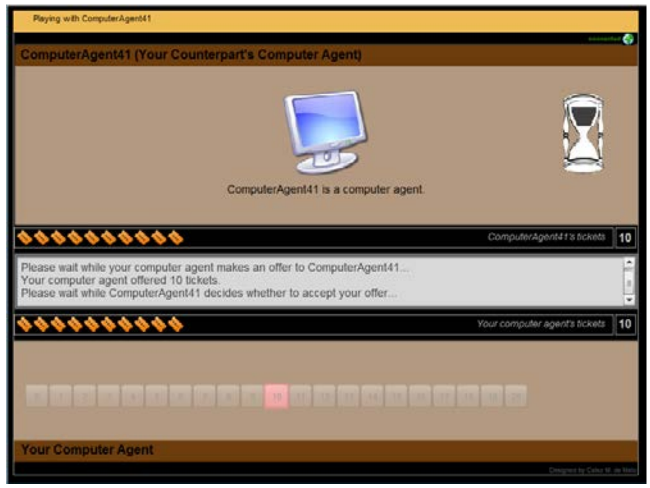

Figure 1. A snapshot of the ultimatum game software.

Responders. Participants were told that the counterparts were either other participants or automated agents that would make decisions on behalf of other participants. Moreover, they were informed that they would play with a different counterpart in each game (i.e., they would play at most once with the same human or agent counterpart). In reality, however, independently of the counterpart type, participants always engaged with the same computer script[2]. To make this manipulation believable, we had people connect to a fictitious server before starting the task for the purposes of "being matched with other participants".

---

[1] The dictator game is another well-known variant of the ultimatum game where the responder always has to accept what the proposer offers and, in this case, isn't even allowed to make a symbolic rejection. The responder, thus, has the least amount of power among the three games. However, since the responder – human or agent – doesn't have to make any decision, we consider the dictator game to be out of scope for our research objectives.

[2] Using this form of deception is not uncommon when studying people's decision making with humans and computers [18], [21]-[23].

Connecting to this server took approximately 30-45 seconds. After concluding the experiment, participants were fully debriefed about this deception.

Proposers. Participants always assumed the role of proposer. They were ostensibly told that this assignment was random. Participants would then interact directly with their counterparts or via an automated agent that would act on their behalf. In the latter case, before starting the task, participants were asked to program their agents to make the offer they wanted.

Sample. We recruited 194 participants from Amazon Mechanical Turk, which is a crowdsourcing platform that allows people to complete online tasks in exchange for pay. Previous research shows that studies performed on Mechanical Turk can yield high-quality data, minimize experimental biases, and successfully replicate the results of behavioral studies performed on traditional pools [55]. We only sampled participants from the United States with an excellent performance history (95% approval rate in previous Mechanical Turk's tasks). Regarding gender, 50.5% of the participants were males. Age distribution was as follows: *22 to 34 years*, 61.3%; *35 to 44 years*, 24.7%; *45 to 54 years*, 9.3%; *55 to 64 years*, 3.6%; *over 65 years*, 1.0%. Professional backgrounds were quite diverse. Participants were paid $2.00 for their participation. Moreover, they had the chance to win extra money, through the lotteries, according to their performance in the tasks. Finally, participants gave their consent before engaging in the experiment and the research presented here was approved by the Internal Review Board at the University of Southern California.

Full anonymity. This experiment was fully anonymous for the participants. To preserve anonymity between participants, human counterparts were referred to as "anonymous" and we never collected any information that could identify the participants. Agents were referred to as "computer agents". To preserve anonymity with respect to the experimenters, we relied on Amazon Mechanical Turk's anonymity system. When interacting with participants from this online pool, researchers are never able to identify the participants, unless they explicitly ask for information that may serve to identify them (e.g., name or photo), which we did not.

Measures. Our main measure was the offers participants made to their counterparts. After completing each task, we had one manipulation check for perception of power: During this event, do you think your counterpart felt powerful (as opposed to powerless)? (1, *Not at all*, to 7, *Very much*). After completing both tasks, we had two further manipulation checks:

- In this experiment, some participants interacted directly with a counterpart, whereas others interacted with computer agents that made decisions on their behalf. In your case, how did you interact with your counterpart? (I interacted directly with my counterpart vs. I interacted via a computer agent that decided on my behalf)
- In this experiment, some participants interacted with another MTurker directly, whereas others interacted with a computer agent that acted on behalf of someone else. In your case, who did you interact with? (I interacted with another MTurker directly vs. I interacted with a computer agent that acted on behalf of someone else).

Finally, to measure social value orientation, before starting the tasks, participants completed the 6-item version of the Slider SVO Measure [56]. After running this measure on our sample, 39.7% of the participants were classified as non-cooperators, and 60.3% as cooperators.

2.2 Results

Manipulation checks. To analyze the manipulation checks for the proposer conditions (Were you interacting via an agent?) and the responder conditions (Were you interacting with an agent?), we ran chi-square tests. The results confirmed that participants accurately remembered the condition they were assigned to (proposers: $\chi^2 = 89.68$, $p < .001$; responders: $\chi^2 = 133.61$, $p < .001$). Regarding the manipulation check for power, participants perceived their counterparts, as expected, to have more power in the ultimatum game ($M = 3.37$, $SE = .134$) than in the impunity game ($M = 2.25$, $SE = .125$), $t(193) = 7.47$, $p < .001$, $r = .473$).

Offers. The offers participants made in the ultimatum and impunity games are shown in Figure 2. To analyze this data, we ran a Responder × Proposer × Power × SVO mixed ANOVA. The results showed a main effect of Responder, with people offering more to humans ($M = 7.13$, $SE = .27$) than to agents representing others ($M = 6.29$, $SE = .26$), $F(1, 186) = 4.98$, $p = .027$, partial $\eta^2 = .026$. This result, thus, supports Hypothesis 1.

The results also reveal a main effect of Proposer, with people offering more when interacting via an agent ($M = 7.09$, $SE = .27$) than when interacting directly with their counterparts ($M = 6.33$, $SE = .26$), $F(1, 186) = 4.18$, $p = .042$, partial $\eta^2 = .022$. This result, therefore, supports Hypothesis 2b (and contradicts Hypothesis 2a).

Regarding power, there was a main effect with people offering more in the ultimatum ($M = 8.27$, $SE = .16$) than in the impunity game ($M = 5.15$, $SE = .30$), $F(1, 186) = 106.00$, $p < .001$, partial $\eta^2 = .363$. However, there were no statistically significant interactions with Responder or Proposer: Power × Responder, $F(1, 186) = .00$, $p = .966$; Power × Proposer, $F(1, 186) = .48$, $p = .489$; Power × Responder × Proposer, $F(1, 186) = 1.51$, $p = .221$.

Finally, regarding social value orientation, there was a main effect with cooperators ($M = 8.01$, $SE = .24$) offering more than non-cooperators ($M = 5.41$, $SE = .29$), $F(1, 186) = 47.86$, $p < .001$, partial $\eta^2 = .205$. There was also a SVO × Power interaction, $F(1, 186) = 12.39$, $p = .001$, partial $\eta^2 = .062$: non-cooperators offered much less in the impunity game than participants in any other condition. There were no statistically significant interactions, however, with Responder or Proposer: SVO × Responder, $F(1, 186) = .08$, $p = .781$; SVO × Proposer, $F(1, 186) = 2.48$, $p = .117$; SVO × Responder × Proposer, $F(1, 186) = .54$, $p = .462$.

3. DISCUSSION

At a time when there is increased interest in agents that act on behalf of humans, this paper raises awareness to important distinctions people still make between interacting with agents that represent humans and interacting with humans directly. Our first main finding is that, everything else being equal, people tend to make more favorable offers to humans than to agents acting on behalf of humans. This result is in line with findings in the behavioral sciences that reveal that perceived social distance

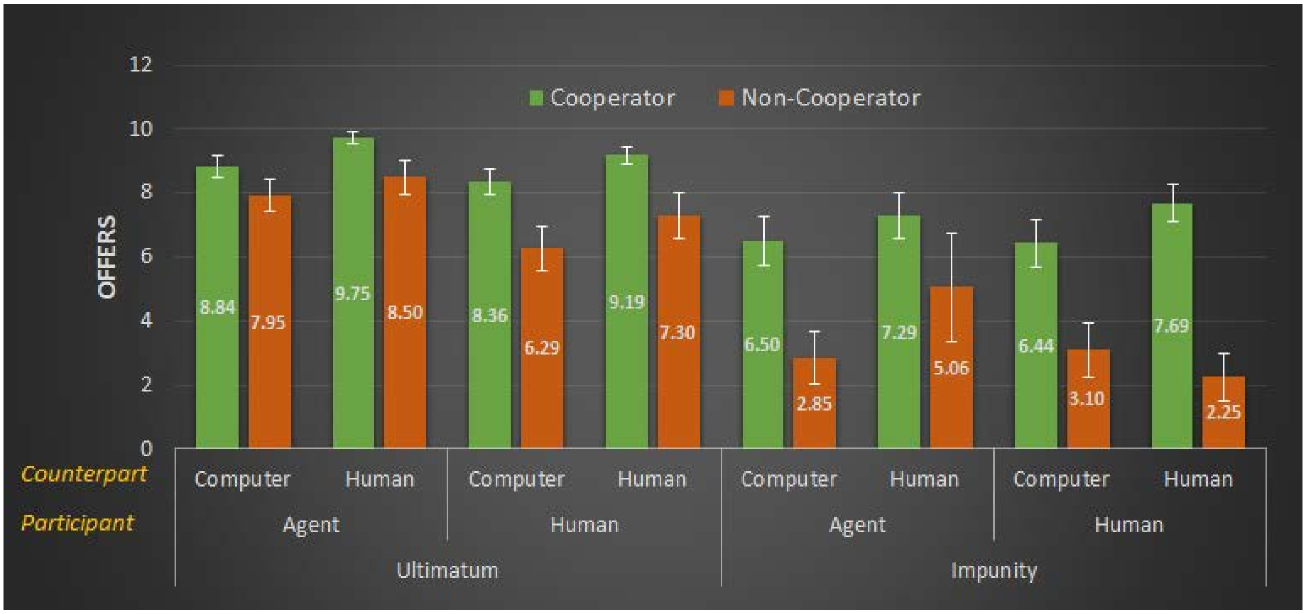

Figure 2. Offers in the ultimatum and impunity games. The error bars show standard errors.

impacts decision making [25]-[31]. Accordingly, the higher the perceived social distance to the counterpart, the less favorable the decisions are likely to be.

Our second main finding is that, in contrast to the tendency to be less fair with agents that represent others, people show higher concern for fairness when tasking an agent to represent them in decision making settings. Despite the fact that acting through an agent increases the social distance with the counterpart, considerations for the longer-term consequences for one's social image seem to prevail [32]-[34]. When creating an agent that is meant to represent one's values, possibly across several interactions with various counterparts, people choose to behave more fairly and preserve their reputations.

A first reading of these findings may suggest that what we accomplished here was to raise a warning to the risk individuals incur of being doubly disadvantaged if they task an agent to interact with other humans. However, instead, we see these findings as introducing two exciting opportunities. First, the results emphasize what needs to be addressed in order to have humans treat agents that represent others as fairly as the humans they are representing, and that is to reduce the perceived social distance to the agents. This can be achieved by emphasizing the presence of the human for whom the agent is working for or, alternatively, by emphasizing shared group membership or common values. In support of this view, previous research demonstrates that people cooperate and trust more with agents that are perceived to share salient physical characteristics (e.g., race [11]) or with which people share a "common fate" (e.g., when they engage in a task as teammates [10], [57]). In fact, by associating agents with membership in multiple positive group memberships, it is even possible to make people favor agents to humans [57].

The second opportunity is that acting through agents can increase a motivation for fairness. The implication is that interaction between agents has the potential to *improve* fairness in society, when compared to the current state-of-affairs in human-human interaction. Because agents do not suffer from the typical constraints we see in humans (e.g., bounded rationality [58]), we already knew that it was possible to use them to increase efficiency in terms of standard economics metrics, such as pareto-optimality [1]-[4]. Here, we propose that agents also have the potential to enhance the kind of social considerations we see in humans [59] – fairness, cooperation, altruism, reciprocity, etc. – by virtue of motivating designers and human users to consider the broader implications of their decisions.

Notice that our proposal is not that people will be more generous when interacting via agents, but rather that they will show higher concerns for fairness. Thus, when playing in the role of proposer, we expect people to make more generous offers than if they were interacting directly with their counterparts; however, when playing in the role of responder, we expect people to be *less* likely to accept unfair offers than if they were not interacting via an automated agent. This is, in fact, an interesting hypothesis to test in future work. These concerns for fairness, for instance, may explain why people show higher demand when interacting via an automated agent than when interacting directly with others in a negotiation task [60].

Our results also demonstrate that power – i.e., the counterpart's ability to shape the final outcome – impacts the decisions we make with automated agents. Our experimental results show that people's offers via agents or with agents were always better when the counterpart was perceived to have more power. This is a replication of earlier findings in the behavioral sciences [41]-[45] and, thus, our contribution is simply to show that the effects of power we see in human-human interaction carry to human-agent interaction.

Additionally, our findings confirm that social value orientation – i.e., people's inherent propensity for cooperation – can also impact behavior with agents. The results showed that cooperators, in general, offered more via agents or with agents than non-cooperators. Given that the literature on social value orientation has mostly focused in social interaction involving exclusively humans [46]-[49], our contribution is in showing that this important individual trait also extends to interactions with automated agents.

The results presented in this paper have important practical implications across several domains. For automated negotiation [1]-[5], the recommendation is that, as discussed above, designers should allow human users to customize their agents. This is likely to lead users to show higher concern for reaching a fair deal. Additionally, designers should strive to minimize the perceived social distance to their human counterparts. These results are also not limited to software agents. As robots get immersed into society [61], [62], the guidelines proposed here for optimizing decision making in human-agent interaction should also be useful for human-robot interaction. Our findings provide insight into how increasing social distance will affect interaction with autonomous vehicles [7] or unmanned flying vehicles [8]: on the one hand, people may react more harshly when something goes wrong because of these vehicles; on the other hand, designers and human will likely strive to embed in them best driving or flying practices. Overall, by reducing social distance with these automated agents and maximizing considerations for the human users' social image, designers can pave the way for more efficient human-agent interaction in various domains.

Finally, this work has interesting ethical implications. There has been considerable concern about allowing these automated agents take their place in society. For instance, the UK's 1998 Data Protection Act gives employees the right to ask for human intervention in the case of any decision made solely by automated means, when personal data is involved. People are also naturally reluctant to let automated vehicles drive on their streets and for unmanned aerial vehicles to apply lethal force [63]. However, experimental work, such as the one presented here, provides critical insight into the psychological mechanisms driving people's behavior with automated agents and, consequently, suggest ways for overcoming these concerns. Moreover, the results in this paper suggest that the very fact that people are forced to think about how they want their agents to behave in society, can lead to a higher-level perspective and, subsequently, to increased motivation for fairness and cooperation.

4. CONCLUSION

In this paper, we show that people's behavior with or via automated agents is different than their behavior with humans. On the one hand, we show that people tend make less favorable offers to agents that act on behalf of humans than if they were interacting with the humans themselves. We propose that to reduce (or increase) this effect we should reduce (or increase) the perceived social distance to the humans on whose behalf the agents are acting. On the other hand, we show that people are likely to demonstrate a higher concern for fairness when interacting via an automated agent than if interacting directly with others. We propose this occurs because people demonstrate a higher concern for their social image and reputation, when tasking an agent to act on their behalf. These results, thus, provide key insight into the psychological mechanisms driving people's behavior with automated agents and, consequently, are key for achieving in human-agent interaction the same kind of efficiency – fairness, cooperation, reciprocity, etc. – we see in human-human interaction.

5. ACKNOWLEDGMENTS

This work is supported by the National Science Foundation, under grant SES-0836004, and the Air Force Office of Scientific Research, under grant FA9550-09-1-0507. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

6. REFERENCES

[1] Lin, R., and Kraus, S. 2010. Can automated agents proficiently negotiate with humans? *Comm. ACM* 53, 78-88.

[2] Jennings, N., Faratin, P., Lomuscio, A., Parsons S., Wooldridge, M., and Sierra, C. 2001. Automated negotiation: Prospects, methods and challenges. *Group Dec. Negot.* 10, 199-215.

[3] Jonker, C., and Robu, V. 2007. An agent architecture for multi-attribute negotiation using incomplete preference information. In *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS'07).*

[4] Sycara, K., and Dai, T. 2010. Agent reasoning in negotiation. In *Handbook of Group Decision and Negotiation*, D. Kilgour, and C. Eden, Eds. Springer Netherlands, 437-451.

[5] Lin, R., Kraus, S., Oshrat, Y., and Gal, Y. 2010. Facilitating the evaluation of automated negotiators using peer designed agents. In *Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI'10).*

[6] Davenport, T., and Harris, J. 2005. Automated decision making comes of age. *MIT Sloan Manage. Rev.* 46, 83-89.

[7] Dresner, K., and Stone, P. 2007. Sharing the road: Autonomous vehicles meet human drivers. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'07).*

[8] Gupte, S. 2012. A survey of quadrotor Unmanned Aerial Vehicles. In *Proceedings of IEEE Southeastcon*. IEEE, 1-6.

[9] Nass, C., and Moon, Y. 2000. Machines and mindlessness: Social responses to computers. *J. Soc. Issues* 56, 81-103.

[10] Nass, C., Fogg, B., and Moon, Y. 1996. Can computers be teammates? *Int. J. Hum-Comput. St.* 45, 669-678.

[11] Nass, C., Isbister, K., and Lee, E.-J. 2000. Truth is beauty: Researching embodied conversational agents. In *Embodied conversational agents*, J. Cassell Ed. MIT Press, Cambridge, MA, 374-402.

[12] Nass, C., Moon, Y., and Carney, P. 1999. Are people polite to computers? Responses to computer-based interviewing systems. *J. App. Psychol.* 29, 1093-1110.

[13] Nass, C., Moon, Y., and Green, N. 1997. Are computers gender-neutral? Gender stereotypic responses to computers. *J. App. Soc. Psychol.* 27, 864-876.

[14] Reeves, B., and Nass, C. 1996. *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

[15] Gray, H., Gray, K., and Wegner, D. 2007. Dimensions of mind perception. *Science* 315, 619.

[16] Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., and Bailenson, J. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 103-124.

[17] Waytz, A., Gray, K., Epley, N., and Wegner, D. 2010. Causes and consequences of mind perception. *Trends Cogn. Sci.* 14, 383-388.

[18] de Melo, C., and Gratch, J. 2015. People show envy, not guilt, when making decisions with machines. In *Proceedings of the Affective Computing and Intelligent Interaction Conference (ACII) 2015*.

[19] Rilling, J., Gutman, D., Zeh, T., Pagnoni, G., Berns, G., and Kilts, C. 2002. A neural basis for social cooperation. *Neuron* 35, 395-405.

[20] Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., and Kircher, T. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLOS ONE* 3, 1-11.

[21] McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proc. Nat. Acad. Sci.* 98, 11832-11835.

[22] Sanfey, A., Rilling, J., Aronson, J., Nystrom, L., and Cohen, J. 2003. The neural basis of economic decision-making in the ultimatum game. *Science* 300, 1755-1758.

[23] Kircher, T., Blümel, I., Marjoram, D., Lataster, T., Krabbendam, L., Weber, J., van Os, J., and Krach, S. 2009. Online mentalising investigated with functional MRI. *Neurosci. Lett.* 454, 176-181.

[24] Gallagher, H., Anthony, J., Roepstorff, A., and Frith, C. 2002. Imaging the intentional stance in a competitive game. *NeuroImage* 16, 814-821.

[25] Hoffman, E., McCabe, K., and Smith, V. 1996. Social distance and other-regarding behavior in dictator games. *Am. Econ. Rev.* 86, 653-660.

[26] Hoffman, E., McCabe, K., Shachat, K., and Smith, V. 1994. Preferences, property rights, and anonymity in bargaining games. *Games Econ. Behav.* 7, 346–380.

[27] Leider, S., Mobius, M., Rosenblat, T., and Do, Q-A. 2009. Directed altruism and enforced reciprocity in social networks. *Q. J. Econ.* 124, 1815-1851.

[28] Goeree, J., McConnell, M., Mitchell, T., Tromp, T., and Yariv, L. 2010. The 1/d law of giving. *Am. Econ. J. Microecon.* 2, 183-203.

[29] Pronin, E., Olivola, C., and Kennedy K. 2008. Doing unto future selves as you would do unto others: Psychological distance and decision making. *Person. Soc. Psychol. Bull.* 34, 224-236.

[30] Nowak, M., and May, R. 1992. Evolutionary games and spatial chaos. *Nature* 359, 826–829.

[31] Nowak, M., Tarnita, C., and Antal T. 2010. Evolutionary dynamics in structured populations. *Philos. Trans. R. Soc. B: Biol. Sci.* 365, 19-30.

[32] Nowak, M., and Sigmund, K. 2005. Evolution of indirect reciprocity. Nature 437, 1291-1298.

[33] Ohtsuki, H. and Iwasa, Y. 2006. The leading eight: social norms that can maintain cooperation by indirect reciprocity. J. Theor. Biol. 239, 435–444.

[34] Andreoni, J. & Bernheim, B. 2009. Social image and the 50–50 norm. A theoretical and experimental analysis of audience effects. Econometrica 77, 1607-1636.

[35] Trope, Y., and Nira, L. 2010. Construal-level theory of psychological distance. *Psychol. Rev.* 117, 440-463.

[36] Agerström, J., and Björklund, F. (2009a). Moral concerns are greater for temporally distant events and are moderated by value strength. *Soc. Cogn.* 27, 261-282.

[37] Agerström, J., and Björklund, F. (2009b). Temporal distance and moral concerns: Future morally questionable behavior is perceived as more wrong and evokes stronger prosocial intentions. *Basic App. Soc. Psychol.* 31, 49-59.

[38] Sanna, L., Chang, E., Parks, C., and Kennedy, L. (2009). Construing collective concerns: Increasing cooperation by broadening construals in social dilemmas. *Psychol. Sci.* 20, 1319-1321.

[39] Henderson, M., Trope, Y., and Carnevale, P. 2006. Negotiation from a near and distant time perspective. *J. Pers. Soc. Psychol.* 91, 712-729.

[40] Giacomantonio, M., De Dreu, C., Shalvi, S., Sligte, D., and Leder, S. 2010. Psychological distance boosts value-behavior correspondence in ultimatum bargaining and integrative negotiation. *J. Exp. Soc. Psychol.* 46, 824-829.

[41] Bacharach, S., and Lawler, E. 1981. *Bargaining: Power, tactics, and outcomes*. San Francisco: Jossey-Bass.

[42] Kelley, H., and Thibaut, J. 1978. *Interpersonal relations: A theory of interdependence*. New York: Academic Press.

[43] Pinkley, R., Neale, M., and Bennett, R. 1994. The impact of alternatives to settlement in dyadic negotiation. *Organ. Behav. Human Dec. Proc.* 57, 97-116.

[44] De Dreu, C. 1995. Coercive power and concession making in bilateral negotiation. *J. Confl. Resolut.* 39, 646-670.

[45] Lawler, E. 1992. Power processes in bargaining. *Sociol. Q.* 33, 17-34.

[46] Messick, D., and McClintock, C. 1968. Motivational bases of choice in experimental games. *J. Exp. Soc. Psychol.* 4, 1-25.

[47] Van Lange, P. 1999. The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *J. Pers. Soc. Psychol.* 77, 337-349.

[48] Bogaert, S., Boone, C., and Declerck, C. 2008. Social value orientation and cooperation in social dilemmas: A review and conceptual model. *Brit. J. Soc. Psychol.* 47, 453-480.

[49] Balliet, D., Parks, C., and Joireman, J. 2009. Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Process. Interg.* 12, 533-547.

[50] Güth, W., Schmittberger, R., and Schwarze, B. 1982. An experimental analysis of ultimatum bargaining. *J. Econ. Behav. & Organ.* 3, 367-388.

[51] J. Henrich, R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. *Am. Econ. Rev*. 91, 73-78.

[52] Camerer, C., and Thaler, R. 1995. Ultimatums, dictators, and manners. *J. Econ. Perspect.*, *9*, 209-219.

[53] Yamagishi, T., Horita, H., Shinada, M., Tanida, S., and Cook, K. 2009. The private rejection of unfair offers and emotional commitment. *Proc. Nat. Acad. Sci.* 106, 11520-11523.

[54] Bolton, G., and Zwick, R. 1995. Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* 10, 95-121.

[55] Paolacci, G., Chandler, J., and Ipeirotis, P. 2010. Running experiments on Amazon Mechanical Turk. *Judgm. Decis. Mak.* 5, 411-419.

[56] Murphy, R., Ackermann, K., and Handgraaf, M. 2011. Measuring social value orientation. *Judgm. Decis. Mak.* 6, 771-781.

[57] de Melo, C., Carnevale, P., and Gratch, J. 2014. Social categorization and cooperation between humans and computers. In *Proceedings of the Annual Meeting of the Cognitive Science Society (CogSci'14).*

[58] Simon, H. 1972. Theories of bounded rationality. In *Decision and Organization*, McGuire, C., and Radner, R., Eds. North-Holland Publishing Company, 161-176.

[59] Rand, D., and Nowak, M. Human cooperation. 2013. *Trends Cogn. Sci.* 17, 413-425. In *Proceedings of the Autonomous Agents and Multi-Agent Systems Conference (AAMAS'04).*

[60] Grosz, B., Kraus, S., and Talman, S. 2004. The influence of social dependencies on decision-making: Initial investigations with a new game. In

[61] Breazeal, C. 2003. Toward sociable robots. *Robotics and autonomous systems* 42: 167-175.

[62] Leite, I., Martinho, C., and Paiva, A. 2013. Social robots for long-term interaction. *Int. J. Soc. Robotic*. 5, 291-308.

[63] Arkin, R. Ethical robots in warfare. *IEEE Technol. Soc. Mag.* 28, 30-33.