The Importance of Cognition and Affect for Artificially Intelligent Decision Makers

Celso M. de Melo[1], Jonathan Gratch[2], and Peter J. Carnevale[1]

[1]USC Marshall School of Business, Los Angeles, CA 90089-0808
demelo@usc.edu, peter.carnevale@marshall.usc.edu

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

Abstract

Agency – the capacity to plan and act – and *experience* – the capacity to sense and feel – are two critical aspects that determine whether people will perceive non-human entities, such as autonomous agents, to have a mind. There is evidence that the absence of either can reduce cooperation. We present an experiment that tests the necessity of both for cooperation with agents. In this experiment we manipulated people's perceptions about the cognitive and affective abilities of agents, when engaging in the ultimatum game. The results indicated that people offered more money to agents that were perceived to make decisions according to their intentions (high agency), rather than randomly (low agency). Additionally, the results showed that people offered more money to agents that expressed emotion (high experience), when compared to agents that did not (low experience). We discuss the implications of this agency-experience theoretical framework for the design of artificially intelligent decision makers.

Introduction

People use different mental processes when making decisions in the presence of others when compared to decisions in non-social settings (Kosfeld et al. 2005; Sanfey et al. 2003). An important factor in social contexts is the attribution of mind to others. Perceptions about mental states provide critical information about others' beliefs, desires, and intentions. This information, in turn, can lead people to cooperate, even when game-theoretic models of rational behavior would predict otherwise (Frank, 2004; Kollock 1998). We argue, therefore, that perceptions of mind in autonomous agents can have a profound impact on people's decision making and is critical for achieving, in human-agent interaction, the kind of efficiency we see in human-human interaction.

Research suggests that people perceive minds in (human and non-human) others according to two dimensions (Gray et al. 2007; Loughnan and Haslam 2007): *agency*, the capacity to plan and act, and *experience*, the capacity to sense and feel. Agency has been a central focus of artificial intelligence research. Considerable attention has been given to achieving optimal rational decisions and, game theory has been successfully applied to develop agents in the context of security, smart grid, traffic management, search-and-rescue, and health applications (Jain, An, & Tambe 2012). However, emotion – a critical component of experience – has comparatively been neglected. Nevertheless, experienced emotion is known to have a deep impact on people's decisions (Bechara, Damasio, and Damasio 2000; Blanchette and Richards 2010; Loewenstein and Lerner 2003) and expressed emotion to influence others' decision making (de Melo et al. 2014; Van Kleef, De Dreu, and Manstead 2010). Thus, designing agents that can achieve with humans the levels of cooperation we see in human-human interaction, requires we consider both cognitive and affective factors (de Melo, Carnevale, and Gratch 2011, 2012; Marsella, Gratch, and Petta 2010; Picard 1997).

A challenge for artificial intelligence, however, is that people tend, by default, to attribute more mind to humans than computers (Blascovich et al. 2002; Gray et al. 2007; Waytz et al. 2010). Effectively, research shows that people tend to reach different decisions, in the same decision tasks for the same financial stakes, with humans when compared to computers. Moreover, people tend to show stronger activation of brain regions associated with emotion and mentalizing (i.e., the inferring of others' mental states) with humans than with computers (Gallagher et al. 2002; Krach et al. 2008; McCabe et al., 2001; Rilling et al. 2002; Sanfey et al. 2003). This is problematic because denial of mind to others can lead to bias against others (Haslam

2006; Loughnan and Haslam, 2007) and subsequent lower levels of trust and cooperation (Sherif et al. 1961).

Building on this agency-experience theoretical framework, we study in this paper how to manipulate perceptions of cognitive and affective mental abilities in agents and its consequence on people's decision making. To accomplish this we present an experiment where people engaged in the ultimatum game with agents that were perceived to possess different mental abilities. In the ultimatum game (Güth, Schmittberger, and Schwarze 1982), one player – the sender – is shown an amount of money and has to decide how much to offer to a second player – the receiver. The receiver then decides whether to accept or reject the offer. If the offer is accepted, then the players get the proposed allocation; however, if the offer is rejected, no one gets anything. The game-theoretical prediction is for the sender to offer the minimum amount (above zero) since the receiver, being rational, will conclude that it is better to have something than nothing. In practice, however, people in industrialized societies tend to offer around 50% of the total amount – i.e., the fair split – and unfair offers (around or below 20% of the total) tend to be rejected 50% of the time (Henrich 2000).

In our experiment, participants were assigned the role of the sender and were informed that the receiver would be an autonomous agent. The agents, however, were manipulated to be perceived as having different cognitive or affective abilities. Regarding cognition, we introduced a variation of the game where the receiver was forced to make a random decision about whether to accept the sender's offer, and the sender knew this ahead of time. This manipulation effectively removed from receivers the ability to act according to their intentions. We compared people's behavior with agents that acted according to their intentions with agents that were forced to make random decisions. Regarding affect, we first gave agents a visual representation consisting of virtual faces (Gratch et al. 2002). Then, we introduced two conditions: in the non-emotional condition (control), agents always showed a neutral display; in the emotional condition, agents expressed positive emotions when the offers were fair and negative emotions when the offers were unfair. Following the predictions of the agency-experience framework, we advanced two hypotheses:

Hypothesis 1 (agency): *People will make more favorable offers to agents that act intentionally, rather than randomly.*

Hypothesis 2 (experience): *People will make more favorable offers to agents that show emotion, rather than those that do not.*

Methods

Design

The experiment followed a 2 × 2 between-participants factorial design: *Intentionality* (random vs. intentional) × *Emotions* (non-emotional vs. emotional). Participants received an initial endowment of 20 tickets and could make an offer between 0 and 20 tickets. Tickets had financial consequences, as they would go into a lottery for a $50 prize in real money. Participants were also informed that agents could win the lottery and if they did then no one else could collect the prize. Agents were described as "computer programs that were designed to decide just like real people" and were labelled "computer agents". Finally, participants engaged in six rounds of the ultimatum game with the same agent.

Regarding the Intentionality factor, intentional agents were described to "make their decisions in whatever way they intended". The actual decision policy, which participants were not told, was based on experimental data on people's behavior (Henrich 2000): if offer was 3 tickets or less, reject; if offer was 4, 5, 6 or 7, accept with 25% chance; if offer was 8 or 9, accept with 75% chance; if offer was 10 or more, accept. Random counterparts were described in the instructions to "throw a die to determine whether they would accept the offer". The decision policy for random counterparts was random (i.e., 50% chance of accepting the offer, independently of the value). Participants were quizzed about these instructions before starting the task; moreover, there were visual reminders in the task pertaining to this factor (e.g., when the agent was deciding, an hour glass would spin for intentional agents but, a rolling die would show for random agents).

Regarding the Emotions factor, participants were given a visual representation that matched their gender. Participants were always matched to agents of the same gender. The emotion policy was as follows: if offer was 3 tickets or less, show anger; if offer was 4, 5, 6, or 7, show sadness; if offer was 8 or more and it was accepted, show joy; if offer was 8 or more and it was rejected, show sadness. This pattern was chosen based on research about the social functions of emotion expressions in decision making (de Melo et al. 2014; Van Kleef et al. 2010) and is compatible with computational approaches for simulating emotion based on appraisal theories (Marsella et al. 2010): if the offer was very unfair, express (an appraisal of) blameworthiness and disapproval of the sender's action through anger; if the offer was moderately unfair, show sadness to supplicate for better offers; if the offer was fair but was rejected (which can occur with random agents), show sadness to appease the sender; finally, if the offer was fair, demonstrate approval by showing joy. The emotion expressions we used are shown in Figure 1. These expressions were validated by de Melo et al. (2014). Figure 2 shows the software implementation.

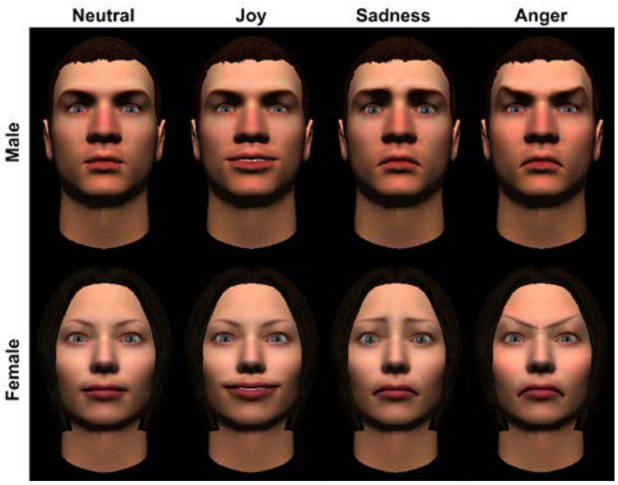

Fig.1 - Facial expressions of emotion.

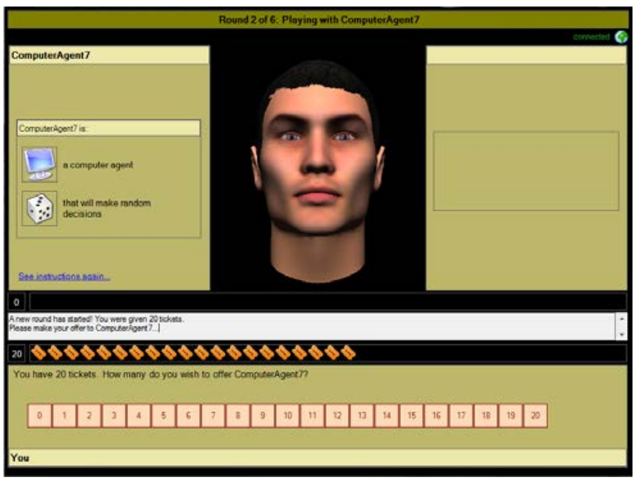

Fig.2 - The software used in the experiment.

Measures

The main dependent variable was the participant's average offer across the rounds. In addition to this behavioral measure, we asked three kinds of classification questions: 1) manipulation checks; 2) perception of mental abilities; 3) subjective impressions of agents. Regarding manipulation checks, after completing the task, we asked the following 7-point classification questions: How were the other party's decisions made? (1, *randomly* to 7, *intentionally*); How much emotion did the other party show? (1, *not at all* to 7, *very much*).

To measure perceptions about counterparts' agency and experience, we asked participants to rate, on a 7-point scale (1, *not at all* to 7, *very much*), the counterpart according to twenty four different items (Gray et al. 2007; Loughnan and Haslam 2007): self-control, morality, planning, thought, cold, consciousness, broadminded, organized, thorough, business-like, automaton-like, acts according to plans and goals, fear, anger, personality, pride, embarrassment, joy, friendly, sociable, jealous, artist-like, warm and acts according to emotions and feelings.

Finally, we asked several exploratory questions regarding the participants' subjective impressions of the agents (scale went from 1, *not at all*, to 7, *very much*): How fair / trustworthy / likable / cooperative was the agent? How likely would you play again with the agent?

Participants and Incentive

We recruited 187 participants at the USC Marshall School of Business. One participant was excluded due to technical error with the software. We ran 12 participants at a time, throughout several sessions. Regarding gender, 57.0% were males. Age distribution was as follows: *21 years and under*, 70.5%; *22 to 34 years*, 29.0%; *35 to 44 years*, 0.5%. All participants were undergraduate or graduate students majoring in Business-related courses and mostly with citizenship from the United States (71.0%). Regarding incentive, first, participants were given school credit for their participation; second, with respect to their goal in the task, participants were instructed to earn as many tickets as possible, as the total amount of tickets would increase their chances of winning the lottery. All participants were asked to give informed consent before starting the task.

Results

Manipulation Checks

We ran Intentionality × Emotions ANOVAs for each manipulation check. Participants perceived random agents' decisions (M = 2.16, SD = 1.88) to be less intentional than intentional agents' decisions (M = 5.24, SD = 1.87), F(1, 182) = 127.92, p = .000, partial η^2 = .413. Participants perceived emotional agents (M = 5.28, SD = 1.67) to show more emotion than non-emotional agents (M = 1.34, SD = .90), F(1, 182) = 396.87, p = .000, partial η^2 = .685. Thus, all manipulation checks passed.

Offers

Since emotion expressions were only shown after the first round, the data analysis focused on rounds 2 to 6. Our formal analysis was based on the average offer across these five rounds. Moreover, to facilitate interpretation, we converted this average into a percentage (over the total amount of 20 tickets). Figure 3 displays the means and 95% confidence intervals for offers in each condition. Table 1 presents further details about this measure.

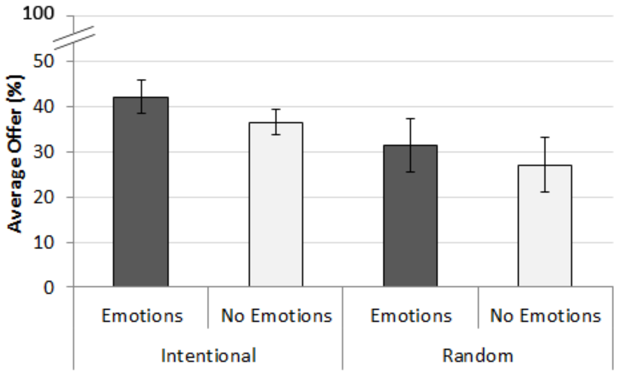

Fig.3 – Means and 95% confidence intervals for participants' offers.

Table 1 – Participants' Average Offers.

		M	*SD*	*95% CI*	
				L	U
Intentional	Emotions	42.11	12.02	38.52	45.69
	No Emotion	36.54	9.33	33.77	39.31
Random	Emotions	31.50	20.37	25.58	37.42
	No Emotion	27.07	20.02	21.05	33.08

Note. Offers are shown as percentages (of the total amount of 20 tickets).

To analyze the data we ran an Intentionality × Emotions ANOVA[1]. This revealed a main effect of Intentionality: people offered more to intentional agents (*M* = 39.36, *SD* = 11.17) than to random agents (*M* = 29.36, *SD* = 20.22), $F(1, 182) = 17.80$, $p = .000$, partial $\eta^2 = .089$, mean difference = 10.04, 95% CI [5.35, 14.74]. There was also a main effect of Emotions: people offered more to emotional (*M* = 36.75, *SD* = 17.57) than non-emotional (*M* = 31.86, *SD* = 16.19) agents, $F(1, 182) = 4.41$, $p = .037$, partial $\eta^2 = .024$, mean difference = 5.00, 95% CI [.30, 9.69]. The Intentionality × Emotions interaction was not statistically significant, $F(1, 182) = .056$, $p = .813$, which suggests intentionality and emotions had an independent and additive effect on participants' offers.

Mind Perception

Principal component analysis (varimax rotation, scree-test) on the agency and experience scale revealed two factors consistent with the literature (Gray et al. 2007; Loughnan and Haslam 2007): agency, explained 36.6% of the variance with main loading factors of planning, thought, thorough, organized and acts according to plans and goals; experience, explained 17.8% of the variance with main loading factors of anger, embarrassment, fear, pride and acts according to emotions and feelings.

To analyze these factors, we ran two separate Intentionality × Emotions ANOVAs. Regarding agency, there was a main effect of Intentionality: people perceived intentional agents (*M* = .24, *SD* = .97) to possess more agency than random agents (*M* = -.36, *SD* = 1.07), $F(1, 182) = 16.14$, $p = .000$, partial $\eta^2 = .081$, mean difference = .60, 95% CI [.31, .90]. The main effect of Emotions and the Intentionality × Emotions interaction were not statistically significant. Regarding experience, there was a main effect of Emotions: people perceived emotional agents (*M* = .27, *SD* = 1.02) to possess more experience than non-emotional agents (*M* = -.46, *SD* = .85), $F(1, 182) = 27.38$, $p = .\ 000$, partial $\eta^2 = .131$, mean difference = .72, 95% CI [.45, 1.00]. The main effect of Intentionality and the Intentionality × Emotions interaction were not statistically significant.

Subjective Impressions

In exploratory fashion, we looked at the participants' subjective impressions of the agents pertaining to fairness, trustworthiness, likability, cooperativeness, and willingness to play again. We ran separate Intentionality × Emotions ANOVAs and we report in this section the main findings for this analysis. Regarding fairness, people perceived intentional agents (*M* = 4.63, *SD* = 1.82) to be fairer than random agents (*M* = 4.05, *SD* = 1.70), $F(1, 182) = 5.11$, $p = .025$, partial $\eta^2 = .027$, mean difference = .58, 95% CI [.07, 1.09]; and, people tended to perceive emotional agents (*M* = 4.56, *SD* = 1.73) to be fairer than non-emotional agents (*M* = 4.12, *SD* = 1.82), $F(1, 182) = 2.96$, $p = .087$, partial $\eta^2 = .016$, mean difference = .44, 95% CI [-.07, .95]. Regarding trustworthiness, people tended to perceive intentional agents (*M* = 4.13, *SD* = 1.78) to be more trustworthy than random agents (*M* = 3.63, *SD* = 1.79), $F(1, 182) = 3.72$, $p = .055$, partial $\eta^2 = .020$, mean difference = .50, 95% CI [-.01, 1.00]; and, people perceived emotional agents (*M* = 4.21, *SD* = 1.77) to be more trustworthy than non-emotional agents (*M* = 3.54, *SD* = 1.76), $F(1, 182) = 6.95$, $p = .009$, partial $\eta^2 = .037$, mean difference = .68, 95% CI [.17, 1.18]. Regarding likability, people liked emotional agents (*M* = 4.09, *SD* = 1.73) more than non-emotional agents (*M* = 3.21, *SD* = 1.70), $F(1, 182) = 12.50$, $p = .001$, partial $\eta^2 = .064$, mean difference = .89, 95% CI [.39, 1.39]. Regarding cooperativeness, people perceived emotional agents (*M* = 4.31, *SD* = 1.73) to be more cooperative than non-emotional agents (*M* = 3.45, *SD* = 1.78), $F(1, 182) = 10.95$, $p = .001$, partial $\eta^2 = .057$, mean difference = .85, 95% CI [.35, 1.36]. Finally, regarding willingness to play again, people were more willing to play

[1] The data meets the assumptions underlying the ANOVA test: (a) the dependent variable can be measured at the interval level; (b) observations in each group were independent, (c) distributions for the data in each group did not differ significantly from normality (Kolmogorov-Smirnov, all *ps* > .05), and (d) Levene's test was significant ($p < .05$) but, the *F*-test in ANOVA has been argued to be robust when the homogeneity of variance assumption is not met and group sample sizes are, as in our case, roughly equal (Glass, Peckham and Sanders, 1972).

again with emotional (M = 4.19, SD = 1.85) than with non-emotional agents (M = 3.51, SD = 1.88), $F(1, 182) = 6.17$, $p = .014$, partial $\eta^2 = .033$, mean difference = .68, 95% CI [.14, 1.23].

Discussion

In this paper we studied how perceptions about autonomous agents' mental abilities influences people's decision making. We build on theory that argues that perceptions of agency (the ability to plan and act) and experience (the ability to sense and feel) influence emergence of cooperation with (human or non-human) others (Gray et al. 2007; Loughnan and Haslam 2007). To test this framework, we described an experiment where participants engaged in the ultimatum game with agents that were perceived to possess different degrees of agency and experience. Confirming Hypothesis 1, the results showed that people made more favorable offers to agents that were perceived to make decisions according to their intentions than agents that made decisions randomly. The results, thus, reinforce the importance of designing agents that possess cognitive abilities to plan and act according to their agency.

The results also showed that people made more favorable offers to agents that displayed positive emotions after fair offers and negative emotions after unfair offers, than agents that displayed no emotion. Therefore, Hypothesis 2 was also supported. This finding is in line with research that emphasizes the social functions of emotion in decision making (de Melo et al. 2014; Van Kleef et al. 2010) and complements existing research that show that emotion expressions can promote cooperation between humans and agents (de Melo et al. 2011, 2012).

The absence of an Intentionality × Emotions interaction suggests that the effects of agency and experience are independent and additive. Further evidence of the distinctiveness of these two factors comes from our results with the perception of agency and experience scales. Principal component analysis of ratings of twenty four items pertaining to various mental characteristics led to two separate factors with loadings compatible with the concepts of agency and experience (Gray et al. 2007; Loughnan and Haslam 2007). Furthermore, the results confirmed that people perceived intentional agents to have more agency than random agents, and emotional agents to possess more experience than non-emotional agents.

Participants also formed different subjective impressions of the agents, according to perceived mental abilities. Intentional agents were perceived to be more fair and trustworthy than random agents. Emotional agents were perceived to be more fair, trustworthy, likable and cooperative than non-emotional agents. Being able to promote favorable impressions of agents can, of course, be invaluable for promoting cooperation in the short- and long-run with humans.

Overall, the results demonstrate that social aspects play a critical role in human-agent interaction. This is in stark contrast to game-theoretic approaches, which fail to capture such social considerations and, thus, are unable to predict the kind of efficiency we see in human-human interaction. For instance, only 18.6% of the participants offered the rational zero amount when engaging with random agents. This percentage was much lower with intentional agents. Moreover, adding emotion to the agents further improved the offers. Indeed, researchers had already argued that the influence of nonverbal behavior, in particular expressions of emotion, is missing in standard economic models of decision making (Boone and Buck 2005; Frank 2004).

These results suggest that perceptions of agency and experience constitute a parsimonious framework to predict how people will make decisions with agents. Earlier research had already shown that people are adept at perceiving agency and causality even in simple shapes that appear to move in a goal-directed manner (Heider and Simmel, 1944). Nevertheless, people still tend to attribute less mind to computers than humans (Blascovich et al. 2002; Gray et al. 2007; Waytz et al. 2010); thus, in line with research that suggests people discriminate when others are perceived to lack mental abilities (Haslam 2006; Loughnan and Haslam 2007), people are likely to show a bias against agents when compared to humans. However, by endowing agents with appropriate cognitive and affective abilities, people become more likely to anthropomorphize the agents (Epley, Waytz, and Cacioppo 2007) and, consequently, treat them in the same manner as humans in a decision making context.

The results, thus, have important implications for the design of agents that engage in decision making with humans. They emphasize that designers need to simulate not only cognitive, but affective capabilities in such agents. Moreover, the results underscore that mere *beliefs* about agency and experience can achieve a powerful effect on people's decision making. This suggests that designers need to be aware of both depth and breadth requirements: on the one hand, it is important to have sophisticated models of rationality and emotion (depth); on the other hand, it is important to have both kinds of models (breadth). When tradeoffs are required, designers should strive to reach the right balance between these qualities. Finally, the results presented in this paper are not only relevant for software agents but, potentially have implications for the design of socially intelligent robots. There is a growing trend in the study of robots that interact naturally with other humans (Breazeal, 2003) and our work reinforces that perceptions about the robots' mental capacities play a key role in determining how people will interact with them.

Regarding future work, it is important to verify whether the findings reported here extend to other decision making contexts. Preliminary results with the iterated prisoner's dilemma (de Melo, Carnevale and Gratch 2013) suggest that some caution is needed when generalizing these findings. In this study, people still cooperated less with agents than with humans, despite both expressing the exact same emotions. However, it is not clear if expression of a different pattern of emotions or the simulation of additional cognitive abilities would suffice to make this distinction go away. Therefore, it is critical we study further manipulations of perceptions of agency and experience in this and other social dilemmas, as well as other decision contexts (e.g., negotiation).

Conclusion

This paper proposes that a framework based on perceptions of agency (the ability to plan and act) and experience (the ability to sense and feel) closes an important gap in our understanding of the way people interact with autonomous agents in decision making. We demonstrated that endowing agents with cognitive and affective mental abilities can increase cooperation in a decision making setting. These results emphasize that it is critical we consider social factors when designing artificially intelligent decision makers, if we wish to promote cooperation with humans and achieve the kind of efficiency we see in human-human interaction.

Acknowledgments

This research was supported in part by the following grants: FA9550-09-1-0507 from the Air Force Office of Scientific Research, IIS-1211064 and SES-0836004 from the National Science Foundation. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

References

Bechara, A., Damasio, H., and Damasio, A. 2000. Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex* 10: 295-307.

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. *Psychological Inquiry* 13: 103-124.

Blanchette, I. and Richards, A. 2010. The influence of affect on higher level cognition: A review of research on interpretation, judgment, decision making and reasoning. *Cognition & Emotion 15*: 1-35.

Boone, R. and Buck, R. 2003. Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior* 27: 163-182.

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

de Melo, C., Carnevale, P., and Gratch, J. 2011. The effect of expression of anger and happiness in computer agents on negotiations with humans. In *Proceedings of Autonomous Agents and Multiagent Systems (AAMAS) 2011*.

de Melo, C., Carnevale, P., and Gratch, J. 2012. The impact of emotion displays in embodied agents on emergence of cooperation with people. *Presence: Teleoperators and Virtual Environments Journal* 20: 449-465.

de Melo, C., Gratch, J., and Carnevale, P. 2013. The effect of agency on the impact of emotion expressions on people's decision making. In *Proceedings of the International Conference of Affective Computing and Intelligent Interaction (ACII).*

de Melo, C., Carnevale, P., Read, S., and Gratch, J. 2014. Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology* 106: 73-88.

Epley, N., Waytz, A., and Cacioppo, J. 2007. On seeing human: A three-factor theory of anthropomorphism. *Psychological Review* 114: 864-886.

Frank, R. 2004. Introducing moral emotions into models of rational choice. In A. S. R. Manstead, N. Frijda & A. Fischer (Eds.), *Feelings and emotions* (pp. 422-440). New York, NY: Cambridge University Press.

Fritz, H. and Simmel, M. 1944. An experimental study of apparent behavior. *The American Journal of Psychology* 57: 243-259.

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

Glass, G., Peckham, P., and Sanders, J. 1972. Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research* 42: 237-288.

Gratch, J., Rickel, J., Andre, E., Badler, N., Cassell, J., and Petajan, E. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems* 17: 54-63.

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

Güth, W., Schmittberger, R., and Schwarze, B. 1982. An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization* 3: 367-288.

Haslam N. 2006. Dehumanization: An integrative review. *Personality and Social Psychology Review* 10: 252-264.

Henrich, J. 2000. Does culture matter in economic behavior? Ultimatum game bargaining among the Machiguenga of the Peruvian Amazon. *The American Economic Review* 90: 973-979.

Jain, M., An, B., and Tambe, M. An overview of recent application trends at the AAMAS conference: Security, sustainability, and safety. *AI Magazine* 33: 14-28.

Kollock, P. 1998. Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology 24*: 183-214.

Kosfeld, M., Heinrichs, M., Zak, P., Fischbacher, U., & Fehr, E. 2005. Oxytocin increases trust in humans. *Nature* 435: 673-676.

Krach, S. et al. 2008. Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE* 3: 1-11.

Loewenstein, G., and Lerner, J. 2003. The role of affect in decision making. In R. Davidson, K. Scherer and H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 619-642). New York, NY: Oxford University Press.

Loughnan, S. and Haslam, N. 2007. Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science* 18: 116-121.

Marsella, S., Gratch, J., and Petta, P. 2010. Computational models of emotion. In K. Scherer., T. Bänziger, and E. Roesch (Eds.), *A blueprint for an affectively competent agent: Cross-fertilization between Emotion Psychology, Affective Neuroscience, and Affective Computing*. New York, NY: Oxford University Press.

McCabe, K., Houser, D., Ryan, L., Smith, V., and Trouard, T. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *PNAS* 98: 11832-11835.

Picard, R. 1997. *Affective computing.* Cambridge, MA: The MIT Press.

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

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.

Sherif, M., Harvey, O., White, B., Hood, W., and Sherif, C. 1961. *Intergroup conflict and cooperation: The Robbers Cave experiment*. Norman, OK: University Book Exchange.

Van Kleef, G., De Dreu, C., and Manstead, A. 2010. An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology* 42: 45-96.

Waytz, A., Gray, K., Epley, N., and Wegner, D. 2010. Causes and consequences of mind perception. *Trends in Cognitive Science* 14: 383-388.