

The Effect of Virtual Agents' Emotion Displays and Appraisals on People's Decision Making in Negotiation

Celso M. de Melo¹, Peter Carnevale², and Jonathan Gratch¹

¹Institute for Creative Technologies, USC, 12015 Waterfront Drive, Building #4 Playa Vista, CA 90094-2536, USA

{demelo,gratch}@ict.usc.edu

²USC Marshall School of Business, Los Angeles, CA 90089-0808, USA
peter.carnevale@marshall.usc.edu

Abstract. There is growing evidence that emotion displays can impact people's decision making in negotiation. However, despite increasing interest in AI and HCI on negotiation as a means to resolve differences between humans and agents, emotion has been largely ignored. We explore how emotion displays in virtual agents impact people's decision making in human-agent negotiation. This paper presents an experiment (N=204) that studies the effects of virtual agents' displays of joy, sadness, anger and guilt on people's decision to counteroffer, accept or drop out from the negotiation, as well as on people's expectations about the agents' decisions. The paper also presents evidence for a mechanism underlying such effects based on appraisal theories of emotion whereby people retrieve, from emotion displays, information about how the agent is appraising the ongoing interaction and, from this information, infer about the agent's intentions and reach decisions themselves. We discuss implications for the design of intelligent virtual agents that can negotiate effectively.

Keywords: Emotion Displays, Decision Making, Negotiation, Appraisal Theories, Reverse Appraisal

1 Introduction

Negotiation, defined by Pruitt and Carnevale [1] as "a discussion among two or more parties aimed at reaching agreement when there is a perceived divergence of interest", is a common mechanism people use for conflict resolution. Negotiation has, thus, drawn considerable attention in the artificial intelligence and human-computer interaction fields as means to resolve conflict in agent-agent and human-agent interaction [2-4]. Automated computer agents can assist less qualified individuals in the negotiation process and, in some situations, even replace human negotiators altogether. Virtual agents, by virtue of having bodies, can add further dimensions to these automated negotiators. One such dimension is the expression of emotions. Effectively, recent decades have seen growing interest on the interpersonal effect of emotion in negotiation (e.g., [5-7]). Complementing research on the impact of emotion in one's own decision making (e.g., [8]), this research explores how one's emotion displays

impact another's decision making and emphasizes that emotional expressions are not simple manifestations of internal experience; rather, expressions are other-directed and communicate intentions, desired courses of actions, expectations and behaviors [9-11]. There is now plenty of empirical evidence demonstrating the differentiated effects of emotions on negotiation outcome (e.g., [5, 12-14]). However, despite the interest artificial intelligence and human-computer interaction has shown in automating negotiation, emotion has been notoriously absent in this endeavor.

To study the social effects of emotion displays in virtual agents on people's decision making in negotiation we build on seminal work by Van Kleef, De Dreu and Manstead on the effects of joy, sadness, anger and guilt on people's concession-making in multi-issue bargaining [12-14]. These studies showed that people conceded more to angry or sad counterparts than happy or guilty counterparts. When facing an angry counterpart, people were argued to infer the other to have high limits and, thus, to avoid costly impasse, were forced to concede; when facing a sad counterpart, people tried to relieve the other's pain by making concessions; when facing a happy counterpart, people inferred the other to have low limits and, thus, could afford to be strategically more demanding; finally, when facing a guilty counterpart, people inferred the other was trying to make amends for a previous transgression and, thus, were more demanding with them. In the past, we have shown that the effects of anger and joy can also occur in human-agent negotiation when emotion is displayed by virtual agents [15]. In the present paper, we complement this work and further study the effects of sadness and guilt in human-agent negotiation. Moreover, in real-life negotiation, people usually can, aside from making counteroffers, accept the counterpart's offer or drop out from the negotiation and it is important to understand when such decisions are made [1]. Therefore, in contrast to the previous study, here we also study how emotion displays impact people's decision to accept or drop out from the negotiation. The paper presents a novel experiment where participants imagined engaging in negotiation with virtual agents that reacted emotionally to the participants' initial offer. We measure how agents' displays of joy, sadness, anger or guilt impact people's decisions to counteroffer, accept or drop out from the negotiation. We also measure how these displays impact people's expectations about the agents' behavior.

This paper also studies the *mechanism* underlying the effects of agents' emotion displays on people's decision making in negotiation. Recently, we proposed a mechanism for the social effects of emotion in the prisoner's dilemma based on appraisal theories of emotion [16]. In appraisal theories [17], emotion displays arise from cognitive appraisal of events with respect to one's goals (e.g., is this event congruent with my goals? Who is responsible for this event?). According to the pattern of appraisals that occurs, different emotions are experienced and displayed. Since displays reflect the agent's intentions through the appraisal process, we argued people could infer, from emotion displays, how virtual agents were appraising the ongoing interaction and, from this information, make inferences about the agents' intentions. This view is compatible with Hareli and Hess' [18] findings that people can, from emotion displays, "reverse engineer" the appraisal process and, from information about appraisals, make inferences about someone's character. This proposal was, thus, referred to as reverse appraisal. Indeed, the results showed that perception of how the agent was

appraising the interaction mediated the effect of emotion displays on perception of the agent’s intention to cooperate in the prisoner’s dilemma. Moreover, in a follow-up experiment, we showed that identical effects could be achieved if, instead of expressing emotions in the face, agents expressed how they were appraising the interaction directly through text (e.g., “I really don’t like this outcome and I blame you for it.”). This result, thus, further supports reverse appraisal. Generalizing these findings, reverse appraisal suggests that what matters for the social effects of emotion in decision making are *not* the displays per se but, the information they communicate about the agent’s appraisals. Applying this to negotiation, reverse appraisal suggests that the effects of displays on people’s decision making in negotiation should also be mediated by people’s perceptions of the agents’ appraisals. Therefore, in this paper we present a multiple mediation analysis [19] that tests whether reverse appraisal can explain the social effects of emotion in human-agent negotiation.

2 Experiment

Scenarios. Participants imagined engaging in a standard multi-issue bargaining task [12] with virtual agents that reacted emotionally to the participant’s initial offer. In this task, the participant played the role of a seller of a consignment of mobile phones whose goal was to negotiate on three issues: price, warranty period and service contract. Each issue had 9 levels, being the highest level the most valuable for the participant, and the lowest level the least valuable. Level 1 on price (\$110) yielded 0 points and level 9 (\$150) yielded 400 points (i.e., each level corresponded to a 50 point increment). Level 1 on warranty (9 months) yielded 0 points and level 9 (1 month) yielded 120 points (i.e., each level corresponded to a 15 point increment). Finally, for duration of service contract, level 1 (9 months) yielded 0 points, and level 9 (1 month) yielded 240 points (i.e., each level corresponded to a 30 point increment). It was pointed out to the participant that the best deal was, thus, 9-9-9 for a total outcome of 760 points ($400 + 120 + 240$). The participant engaged in this task with a virtual agent, referred to as David, which had different payoffs that were not known. The negotiation supposedly proceeded according to the alternating offers protocol, being the participant the first to make an offer. The participant was informed that the negotiation would proceed until one player *accepted* the other’s offer, *dropped out* from the negotiation or time expired. In reality, the scenario pertained only to the first offer made by the participant which corresponded to 9-6-7, worth $400 + 75 + 180 = 655$ points to the participant. After being shown the offer, the participant watched a video of the agent reacting emotionally. After watching the video, the participant was asked several questions and, then, explained that the scenario was over.

Conditions. The experiment had one between-participants factor: Emotion Display (Neutral vs. Joy vs. Sadness vs. Anger vs. Guilt). The emotion displays were shown in the face of a virtual agent and the respective displays are shown in Figure 1. These facial displays were animated using a muscular model of the face with blushing and wrinkles [20].

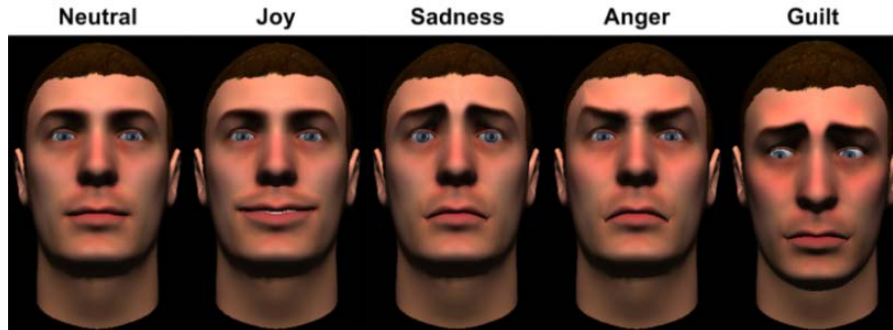


Fig. 1. The facial displays of emotion.

Measures. After watching the video of the agent’s emotional reaction, we asked participants the following questions: How much did David experience each of the following emotions a) Sadness b) Joy c) Anger d) Guilt? (scale went from 1, *not at all*, to 7, *very much*).

Even though several appraisal theories have been proposed [17, 21-23], there tends to be agreement on which appraisals predict the emotions we considered in this experiment: joy occurs when the event is conducive to one’s goals; sadness occurs when the event is not conducive to one’s goals; anger occurs when the event is not conducive to one’s goals and is caused by another agent; guilt occurs when the event is not conducive to one’s goals and is caused by the self. Thus, two appraisal variables are of relevance here: (a) *conduciveness to goals*, which measures whether the event is consistent or inconsistent with the individual’s goals; and, (b) *blameworthiness*, which measures whether the self or another agent is responsible for the event. After watching the video of the agent’s emotional reaction, participants were asked the following questions about how was the agent appraising the outcome (all questions on a scale from 1, *not at all*, to 7, *very much*):

1. How pleasant for David was it to be in this situation? [21]
2. At the time of experiencing the emotion, do you think David perceived that the consequences of the event did or would bring about positive, desirable consequences for him (e.g., helping him reach a goal, or giving pleasure)? [22]
3. Was the situation obstructive or conducive to David’s goals? [23]
4. Was what happened something that David regarded as fair? [23]
5. How much did you think David blamed himself for the event? [21]
6. How much did you think David blamed you for the event? [21]

Following the appraisal perception questions, we asked several questions about the participant’s decision and expectations regarding the agent’s decision (Questions 7, 8, 11 and 12 were on a scale from 1, *not at all*, to 7, *very much*):

7. How likely is David to accept your offer?
8. How likely is David to drop out from the negotiation?
9. If David were to make a counter-offer, what would that likely be?
10. How likely are you to accept David’s counter-offer?

11. How likely are you to drop out from the negotiation?
12. If you were to make a counter-offer what would that be?

Finally, two questions were asked that characterized the interaction with the agent (scale went from 1, *not at all*, to 7, *very much*):

13. How cooperative was David?
14. Would you negotiate again with David in a future occasion?

Participants. We recruited 204 participants online using Amazon Mechanical Turk. This resulted in approximately 41 participants for each emotion. Regarding gender, 56.4% of the participants were male. Age distribution was as follows: *18 to 21 years*, 16.2%; *22 to 34 years*, 56.9%; *35 to 44 years*, 15.7%; *45 to 54 years*, 7.4%; *55 and over*, 3.9%. Most participants were from the United States (40.7%) and India (42.2%). The education level distribution was as follows: *high school*, 18.6%; *college*, 57.8%; *Masters*, 19.6%; *Ph.D. or above*, 3.9%. Education majors and profession were quite diverse. Participants were paid USD \$1.02 and average participation time was 15 minutes.

3 Results

Manipulation Check. To validate that participants took the task seriously we looked at outliers and performance measures (e.g., participation time). No participants were excluded under these criteria. To validate that participants were interpreting the virtual agent's facial displays as intended, we subjected the emotion interpretation measures to a MANOVA across Emotion Display. Table 1 shows the means and standard deviations for this analysis. The results for the multivariate test showed a significant effect ($p < .05$, Pillai's Trace). ANOVAs confirmed significant differences for the emotion interpretation measures (see last column of Table 1), and subsequent Bonferroni post-hoc tests revealed that: perception of joy was highest for Joy, perception of sadness was highest for Sadness, perception of anger was highest for Anger and, perception of guilt was highest for Guilt (all tests, $p < .05$). Thus, the agents' emotion displays were being interpreted as intended.

Table 1. Perceived emotion in the facial displays (manipulation check)

Perceived Emotion	Emotion Display					ANOVA
	Neutral	Joy	Sadness	Anger	Guilt	
Joy	2.20 (1.38)	5.71 (1.08)	1.95 (1.28)	2.10 (1.53)	1.98 (1.07)	66.63*
Sadness	2.00 (1.32)	1.34 (0.88)	5.27 (1.36)	3.32 (1.89)	4.53 (1.65)	52.25*
Anger	1.93 (1.39)	1.34 (0.86)	2.41 (1.50)	4.51 (1.38)	2.48 (1.18)	35.78*
Guilt	1.93 (1.23)	1.24 (0.66)	2.59 (1.52)	2.76 (1.92)	3.78 (1.51)	17.95*

* $p < .0125$ (Bonferroni correction applied).

Effects on Decision Measures. To study the participant's counteroffer and the agent's expected counteroffer we looked at *demand difference*, a standard measure that is defined as the difference between demand (i.e., the value of the offer in points) in the participant's first offer (655 points) and the counteroffer. The means and standard errors for demand difference are shown in Figure 2. The means and standard errors for the remaining decision measures and agent characterization measures are shown in Figure 3. Table 2 summarizes descriptive statistics for all these measures. To analyze these results we ran one-way ANOVAs, across Emotion Display, for each measure. Tukey HSD post-hoc tests were used for pairwise comparisons.

Regarding participant's demand difference, there was a (main) effect of Emotion Display ($F(4,199) = 2.85, p < .05$) and post-hoc tests revealed that demand difference was lower after the agent displayed Joy than Anger ($p < .05$) or Guilt ($p < .10$). Regarding (expected) agent's demand difference, there was an effect of Emotion Display ($F(4,199) = 6.02, p < .05$) and post-hoc tests revealed that expected demand difference was lower after the agent displayed Joy than Anger ($p < .05$), Guilt ($p < .05$) or Sadness ($p < .10$).

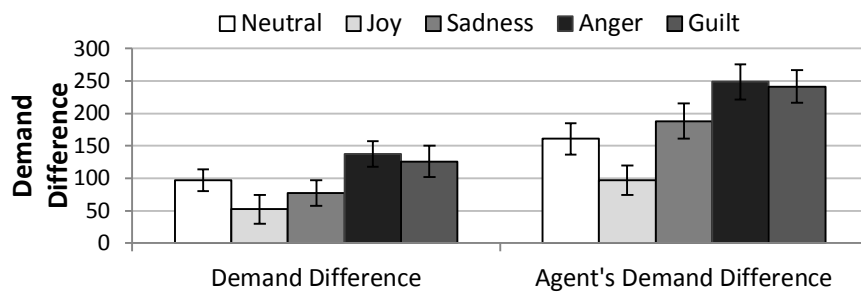


Fig. 2. Means (and standard errors) for demand difference (difference in points between the participant's first offer and the counteroffer).

Regarding the participant's decision to accept, there was a (main) effect of Emotion Display ($F(4,199) = 5.47, p < .05$) and post-hoc tests revealed the participant was more likely to accept after Joy than Sadness, Anger or Guilt (all tests, $p < .05$). Regarding the (expected) agent's decision to accept, there was an effect of Emotion Display ($F(4,199) = 48.19, p < .05$) and post-hoc tests revealed that the agent was more likely to accept after displaying Joy than the Neutral display and, less likely to accept after displaying Sadness, Anger or Guilt than the Neutral display (all tests, $p < .05$). Regarding the participant's decision to drop out, there was an effect of Emotion Display ($F(4,199) = 2.86, p < .05$) and post-hoc tests revealed the participant was more likely to drop out after the a display of Sadness than Joy ($p < .05$). Regarding the (expected) agent's decision to drop out, there was an effect of Emotion Display ($F(4,199) = 19.52, p < .05$) and post-hoc tests revealed that the agent was more likely to drop out after displaying Sadness, Anger or Guilt than the Neutral display and, less likely to drop out after displaying Joy than the Neutral display (all tests, $p < .05$).

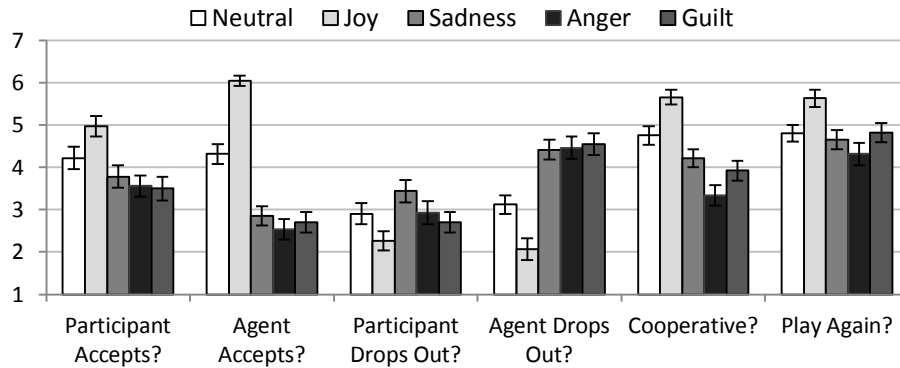


Fig. 3. Means (and standard errors) for accept, drop out and characterization measures.

Table 2. Means (and standard deviations) for decision measures

	Neutral	Joy	Sadness	Anger	Guilt
Participant's Demand	96.95	52.20	77.07	137.56	126.13
Difference *	(109.15)	(143.68)	(126.63)	(125.21)	(152.96)
Agent's Demand Difference *	160.98	96.95	188.17	248.90	241.63
	(154.35)	(147.02)	(177.96)	(173.10)	(158.78)
Participant Accepts *	4.22	4.98	3.78	3.56	3.50
	(1.71)	(1.56)	(1.71)	(1.60)	(1.75)
Agent Accepts *	4.32	6.05	2.85	2.54	2.70
	(1.51)	(0.77)	(1.42)	(1.53)	(1.51)
Participant Drops Out *	2.90	2.27	3.44	2.93	2.70
	(1.61)	(1.43)	(1.66)	(1.74)	(1.54)
Agent Drops Out *	3.12	2.07	4.41	4.46	4.55
	(1.44)	(1.65)	(1.50)	(1.67)	(1.65)
Play Again *	4.80	5.63	4.66	4.32	4.83
	(1.29)	(1.32)	(1.46)	(1.66)	(1.41)
Cooperative *	4.76	5.66	4.22	3.34	3.93
	(1.41)	(1.13)	(1.35)	(1.56)	(1.49)

Note. Demand difference is the difference in points between the participant's first offer (9-6-7) and the counteroffer.

* $p < .05$.

Finally, Regarding the participant's willingness to play again with the agent, there was a (main) effect of Emotion Display ($F(4,199) = 4.67, p < .05$) and post-hoc tests revealed that the participant was more likely to play again with the agent that displayed Joy than Neutral ($p < .10$), Sadness ($p < .05$), Anger ($p < .05$) and Guilt ($p < .10$). Regarding perception of how cooperative the agent was, there was an effect of Emotion Display ($F(4,199) = 16.21, p < .05$) and post-hoc tests revealed that the agent that displayed Joy was perceived as more cooperative than the agent that showed the

Neutral display ($p < .05$) and the agent that showed Anger ($p < .05$) or Guilt ($p < .10$) was perceived to be less cooperative than the agent that showed the Neutral display.

Effects on Perception of Appraisals Measures. Questions 1 to 4 were highly correlated ($\alpha = .903$) and, thus, were collapsed (averaged) into a single measure called conduciveness to goals. The means, standard deviations and standard errors for conduciveness to goals, self-blame (Question 5) and participant-blame (Question 6) are displayed in Figure 4 and presented in Table 3. For our main analysis, we subjected the perception of appraisal measures to a one-way ANOVA across Emotion Display. Tukey HSD post-hoc tests were used for pairwise comparisons.

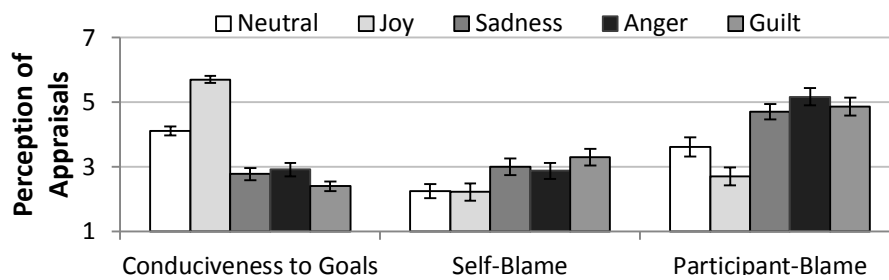


Fig. 4. Means (and standard errors) for perception of appraisals.

Table 3. Means (and standard deviations) for perception of appraisals

	Neutral	Joy	Sadness	Anger	Guilt
Conduciveness to Goals *	4.11 (0.92)	5.71 (0.68)	2.77 (1.26)	2.92 (1.33)	2.40 (0.97)
Self-Blame *	2.24 (1.43)	2.22 (1.68)	3.00 (1.63)	2.88 (1.58)	3.30 (1.68)
Participant-Blame *	3.61 (1.90)	2.71 (1.76)	4.71 (1.52)	5.17 (1.67)	4.88 (1.76)

* $p < .05$.

Regarding conduciveness to goals, the results showed a (main) effect of Emotion Display ($F(4, 199) = 66.29, p < .05$). Post-hoc tests revealed that when the agent displayed Joy, it was perceived to find the outcome more conducive than when it showed the Neutral display; moreover, when the agent displayed Sadness, Anger or Guilt, the agent was perceived to find the outcome less conducive than when it showed the Neutral display (all tests, $p < .05$). Regarding self-blame, the results showed an effect of Emotion Display ($F(4, 199) = 3.62, p < .05$). Post-hoc tests revealed that when the agent displayed Guilt it was perceived to blame itself more than when it showed Joy or the Neutral Display (all tests, $p < .05$). Regarding participant-blame, the results showed an effect of Emotion Display ($F(4, 199) = 14.52, p < .05$). Post-hoc tests revealed that when the agent displayed Joy it was perceived to blame the participant the least and, when it showed Sadness, Anger or Guilt it was perceived to blame the participant more than when it displayed the Neutral display (all tests, $p < .05$).

Mediation Analysis. Here we present a causal steps approach multiple mediation analysis [19] of perceptions of appraisal on the effect of emotion displays on participant's demand difference and (expected) agent's demand difference. This method is an extension to multiple mediators of the single-mediation analysis proposed by Baron and Kenny [24]. Figure 5 summarizes the mediation model. The independent variables (IVs) were the classification questions for perception of joy, sadness, anger and guilt. The dependent variables (DV) were the demand difference measures. The proposed mediators were the perception of appraisal variables: conduciveness to goals, self-blame and participant-blame. According to this approach, there is mediation by a specific mediator M_x if: (1) the path, a_x , from the IV to the mediator is significant; (2) the path, b_x , from the mediator to the DV, when controlling for the IV, is significant; (3) the indirect effect, $a_x b_x$, from the IV to the DV, when controlling for the mediator, is significantly different than zero and greater than zero by a non-trivial amount. Moreover, there is mediation of the *set* of mediators when the sum of the indirect effects of all mediators is significantly different than zero. There is full mediation when the total effect, c , from the IV to the DV (not considering any mediators), becomes non-significant, i.e., the direct effect, c' , from the IV to the DV (when accounting for all mediators), is non-significant. Finally, in the original paper, Baron and Kenny also require the total effect, c , to be significant. However, many authors advocate this path need not be significant, in the multiple mediation case, for mediation to occur [19].

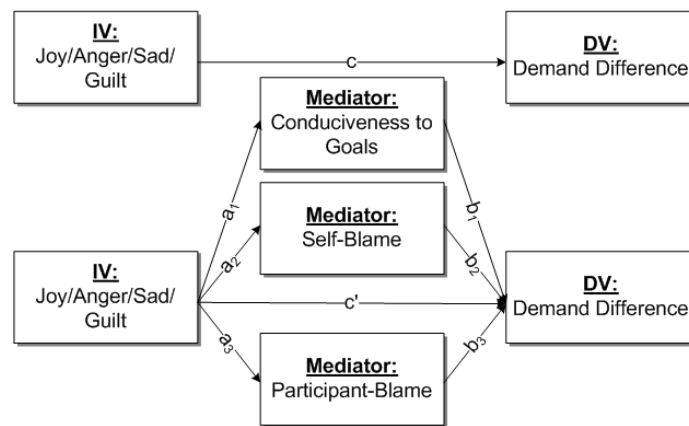


Fig. 5. The multiple mediation model.

Table 4 shows the analysis: the shaded cells on the a , b and ab path columns represent that the causal-step requirement on the respective path has been passed. Regarding the participant's demand difference, the results showed full mediation of conduciveness to goals on the effect of Joy, Sadness and Anger; there was also evidence of mediation of conduciveness to goals and self-blame on the effect of Guilt. Regarding (expected) agent's demand difference, the results showed full mediation of conduciveness to goals and participant-blame on the effect of Joy, Sadness and Anger; there was also evidence of mediation of conduciveness to goals on the effect of Guilt.

Table 4. Mediation analysis of perceptions of appraisals on the effect of emotion displays

	IV→ Mediators (<i>a paths</i>)			Mediators → DV (<i>b paths</i>)			Total Effect (<i>c path</i>)	Direct Effect (<i>c' path</i>)	Indirect Effect (<i>ab paths</i>)				
	Cn	SB	PB	Cn	SB	PB			Tot	Cn	SB	PB	
Participant's Demand Difference	Joy	.62** (.000)	-.01 (.829)	-.36** (.000)	-21.15** (.004)	-5.03 (.227)	-1.70 (.670)	-9.42** (.008)	3.20 (.567)	-12.62** (.004)	-13.36** (.003)	.21 (.440)	.53 (.669)
	Sad	-.39** (.000)	.22** (.000)	.29** (.000)	-17.11** (.001)	-5.65 (.198)	-1.81 (.651)	6.08* (.066)	1.69 (.665)	4.39** (.048)	6.11** (.002)	-1.32 (.204)	-.41 (.650)
	Anger	-.29** (.000)	.12* (.077)	.42** (.000)	-17.01** (.001)	-5.67 (.178)	-2.62 (.521)	7.58* (.052)	4.28 (.310)	3.29* (.077)	4.95** (.002)	-.74 (.220)	-.92 (.521)
	Guilt	-.19** (.005)	.55** (.000)	.14* (.097)	-17.21** (.000)	-7.7* (.099)	-1.82 (.648)	4.56 (.264)	5.31 (.256)	-.75 (.766)	3.03** (.008)	-3.60* (.098)	-.18 (.656)
Agent's Demand Difference	Joy	.62** (.000)	-.01 (.829)	-.36** (.000)	-32.95** (.004)	-11.72* (.074)	12.21* (.055)	-31.93** (.000)	-7.16 (.400)	-24.77** (.000)	-20.55** (.004)	.15 (.830)	-4.37* (.066)
	Sad	-.39** (.000)	.22** (.000)	.29** (.000)	-44.53** (.000)	-9.60 (.157)	12.71** (.046)	10.93* (.059)	-7.98 (.201)	18.91** (.000)	17.27** (.000)	-2.06 (.177)	3.69* (.063)
	Anger	-.29** (.000)	.12* (.077)	.42** (.000)	-39.74** (.000)	-12.09* (.067)	11.85* (.072)	15.83** (.028)	.93 (.895)	14.90** (.000)	11.37** (.001)	-1.49 (.198)	5.01* (.083)
	Guilt	-.19** (.005)	.55** (.000)	.14* (.097)	-41.14** (.000)	-7.79 (.319)	12.13* (.056)	-2.60 (.723)	-7.77 (.330)	5.17 (.363)	7.81** (.013)	-4.31 (.319)	1.67 (.206)

Note. Cn = Conduciveness to goals; SB = Self-Blame; PB = Participant-Blame; CP = Coping Potential.

Demand difference is the difference in points between the participant's first offer (9-6-7) and the counteroffer.

Values correspond to unstandardized regression coefficients (*p* values in parentheses).

** *p* < .05. * *p* < .10.

4 Discussion

The results showed that emotion displays in virtual agents can impact people's decision making in negotiation. In line with previous findings [12, 15], our results indicated that people conceded more to an angry agent than to a happy agent. The rationale here is that people infer the angry agent to have high aspirations and, thus, to avoid costly impasse are forced to concede; in contrast, people infer the happy agent to have low aspirations and, therefore, can afford to be strategically more demanding. However, our results suggested displays of guilt led people to concede more. This finding seemingly contrasts with Van Kleef et al.'s findings [13] that suggest people concede less to a guilty counterpart because guilt is signaling an apology and a willingness to make amends for a previous transgression. However, our scenarios pertained to an emotional reaction to the participant's first offer and, therefore, there is no obvious previous transgression. In this case, participants are more likely interpreting the display of guilt as serving a supplication function (i.e., a cry for help) and, therefore, comply by helping the agent by making more generous offers. These different interpretations of guilt reinforce the importance of context for the interpretation of emotion and the social effects of emotion displays [6, 14, 25].

Regarding the decisions to accept or drop out, the results showed that people were more likely to accept and expect the agent to accept after a display of joy than a display of sadness, anger or guilt; complementary, people were more likely to drop out from the negotiation or expect the agent to drop out after a display of sadness, anger or guilt than a display of joy. Regarding characterization, the results indicated people, if given a choice, would likely negotiate again with the smiling agent, which they found to be the most cooperative, and would likely not negotiate again with the angry agent, which they found to be the least cooperative.

To understand the mechanism behind such effects, we looked at the reverse appraisal proposal [16, 18] whereby people retrieve, from emotion displays, information about how agents are appraising the ongoing interaction and, from this information, infer the agents' intentions. Effectively, our results confirmed that people were able to retrieve, from emotion displays, information about agent's appraisals: the offer was perceived to be more conducive to the agent's goals after a display of joy than a display of sadness, anger or guilt; the agent was perceived to blame more itself after a display of guilt; and, the agent was perceived to blame the participant more after displaying anger, guilt or sadness than joy. These results are, in general, compatible with expectations from appraisal theories [17] and with previous findings that show people can retrieve information about appraisals from facial displays of emotion [26]. Moreover, our results show that perceptions of the agents' appraisals mediated the effects of emotion displays on people's decisions and expectations about the agents' decisions. For instance, the effect of joy or anger on people's demand was fully mediated by perceptions about how conducive the offer was to the agent. This is compatible with Van Kleef et al.'s argument that people use displays of joy or anger to infer the counterpart's limits in negotiation [12]. These results, thus, suggest that information

about the agents' appraisals are a critical component people retrieve from emotion displays to inform their decision making in negotiation.

Overall, the results emphasize the importance of valence for the social effects of emotion displays in human-agent negotiation. Whereas it is easy to distinguish between the effects of joy and the negative displays, the differences between the effects of sadness, anger and guilt are more subtle. The mediation analysis also lends some support to this dichotomy in that conduciveness to goals, which can be viewed as valence, is the most significant mediator for the social effects of emotion. However, it is important not to over-generalize these findings and conclude that valence is all that matters. Effectively, there is consistent evidence in the literature for the differentiated effects of negative displays of emotion on people's decision making (for a review, see [14]). Our own research has shown that sadness, anger, and guilt can have distinct effects on people's decision to cooperate in the prisoner's dilemma and, that these effects are mediated by, not only conduciveness to goals, but other appraisals as well [16]. Finally, though currently lacking, research on displays of positive emotions (joy, pride, admiration, gratitude, etc.) is likely to reveal, as well, differentiated social effects on people's decision making.

Regarding limitations and future work, a methodology based in scenarios can always be subjected to the criticism that results only tap into people's naïve theories of the social effects of emotion and not actual behavior [27]. In this regard, we point out that the results presented here are compatible with other findings where people actually engaged in negotiation with emotional human [12, 13] and virtual [15] counterparts. Still, we plan to replicate and extend this work with experiments, with proper financial incentives [28], where people engage in negotiation with emotional virtual agents. Furthermore, these experiments should pay special attention to the contextual effects of emotion. As discussed above, the interpretation of emotion displays is highly contextual [6, 14, 25]; for instance, a display of guilt after the human's first offer can have a very different meaning than the same display of guilt after the agent has made an unreasonable counter-offer.

Finally, the results presented in this paper have important implications for the design of agents that can negotiate. First, the results report which emotions agents should express to systematically influence humans' concession-making, likelihood of accepting or dropping out and, perceptions of the agent's cooperativeness. Second, the results also emphasize tradeoffs designers need to be aware of when endowing agents with emotions: Whereas an agent that displays anger, guilt or sadness might be effective in eliciting concessions from people, people are less willing to interact with such agents in the future; on the other hand, whereas people might prefer to interact with an agent that displays joy, the agent might not be very successful in reaching profitable agreements for itself, at least in the short run. Finally, the mediation results lend support to reverse appraisal which suggests that what is essential for the social effects of emotion in negotiation are not the emotion displays per se but, the information about appraisals that is communicated. This emphasizes that designers need not necessarily simulate facial displays of emotion to achieve the effects of emotion on negotiation we see in people; all that is required is to effectively convey the underlying appraisals.

Acknowledgments

This work was sponsored by the Fundação para a Ciência e a Tecnologia grant #SFRH-BD-39590-2007; the Air Force Office of Scientific Research under grant FA9550-09-1-0507; and, the NSF under grant IIS-0916858. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

1. Pruitt, D., & Carnevale, P. (1993). *Negotiation in social conflict*. Pacific Grove, CA: Brooks/Cole Publishing Co.
2. Jennings, N., Faratin, P., Lomuscio, A., Parsons S., Wooldridge, M., & Sierra, C. (2001). Automated negotiation: Prospects, methods and challenges. *Group Decis. Negot.*, 10(2), 199-215.
3. Faratin, P., Sierra, C., and Jennings, N. 2002. Using similarity criteria to make issue trade-offs in automated negotiations. *Artificial Intelligence*, 142(2), 205–237.
4. Lin, R., & Kraus, S. (2010). Can automated agents proficiently negotiate with humans? *Communications of the ACM*, 53(1), 78-88.
5. Van Kleef, G., De Dreu, C., & Manstead, A. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Adv. Exp. Soc. Psychol.*, 42, 45-96.
6. Hareli, S., & Hess, U. (2012). The social signal value of emotions. *Cognition & Emotion*, 26(3), 285-289.
7. Morris, M., & Keltner, D. (2000). How emotions work: An analysis of the social functions of emotional expression in negotiations. *Res. Organ. Behav.*, 22, 1-50.
8. Blanchette, I., & 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.
9. Frijda, N., & Mesquita, B. (1994). The social roles and functions of emotions. In S. Kitayama & H. Markus (Eds.), *Emotion and culture: Empirical studies of mutual influence* (pp. 51–87). Washington, DC: American Psychological Association.
10. Keltner, D., & Haidt, J. (1999). Social functions of emotions at four levels of analysis. *Cognition & Emotion*, 13(5), 505–521.
11. Keltner, D., & Kring, A. M. (1998). Emotion, social function, and psychopathology. *Rev. Gen. Psychol.*, 2, 320–342.
12. Van Kleef, G., De Dreu, C., & Manstead, A. (2004). The interpersonal effects of anger and happiness in negotiations. *J. Pers. Soc. Psychol.*, 86, 57-76.
13. Van Kleef, G., De Dreu, C., & Manstead, A. (2006). Supplication and appeasement in negotiation: The interpersonal effects of disappointment, worry, guilt, and regret. *J. Pers. Soc. Psychol.*, 91, 124-142.
14. Van Kleef, G., De Dreu, C., & Manstead, A. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, 42(10), 45-96.
15. de Melo, C., Carnevale, P., & 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*.

16. de Melo, C., Carnevale, P., Read, S., & Gratch, J. (2012). Reverse appraisal: The importance of appraisals for the effect of emotion displays on people's decision-making in a social dilemma. To be presented at *The 34th Annual Meeting of the Cognitive Science Society (CogSci) 2012*.
17. Ellsworth, P., & Scherer, K. (2003). Appraisal processes in emotion. In R. Davidson, K. Scherer, H. Goldsmith (Eds.), *Handbook of Affective Sciences* (pp. 572-595). New York, NY: Oxford University Press.
18. Hareli, S., & Hess, U. (2010). What emotional reactions can tell us about the nature of others: An appraisal perspective on person perception. *Cognition & Emotion*, *24*, 128-140.
19. Preacher, K., & Hayes, A. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Meth.*, *40*, 879-891.
20. de Melo, C., & Gratch, J. (2009b). Expression of emotions using wrinkles, blushing, sweating and tears. In *Proceedings of the IVA 2009*, pp. 188-200.
21. Smith, C., & Ellsworth, P. (1985). Patterns of cognitive appraisal in emotion. *J. Pers. Soc. Psychol.*, *48*, 813-838.
22. Scherer, K. (2001). Appraisal considered as a process of multi-level sequential checking. In K. Scherer, A. Schorr & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 92-120). New York, NY: Oxford University Press.
23. Frijda, N. (1989). Relations among emotion, appraisal, and emotional action readiness. *J. Pers. Soc. Psychol.*, *57*, 212-228.
24. Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *J. Pers. Soc. Psychol.*, *51*, 1173-1182.
25. Aviezer, H., Hassin, R., Ryan, J., Grady, C., Susskind, J., & Anderson, A. et al. (2008). Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological Science*, *19*(7), 724-732.
26. Scherer, K., & Grandjean, D. (2008). Facial expressions allow inference of both emotions and their components. *Cognition & Emotion*, *22*(5), 789-801.
27. Parkinson, B., & Manstead, A. S. R. (1993). Making sense of emotion in stories and social life. *Cognition & Emotion*, *7*(3/4), 295-323.
28. Hertwig, R., & Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behav. Brain Sci.*, *24*(3), 383-451.