Comparing Behavior Towards Humans and Virtual Humans in a Social Dilemma

Rens Hoegen(✉), Giota Stratou, Gale M. Lucas, and Jonathan Gratch

Institute for Creative Technologies, University of Southern California, Los Angeles, USA
{rhoegen,stratou,lucas,gratch}@ict.usc.edu

Abstract. The difference of shown social behavior towards virtual humans and real humans has been subject to much research. Many of these studies compare virtual humans (VH) that are presented as either virtual agents controlled by a computer or as avatars controlled by real humans. In this study we directly compare VHs with real humans. Participants played an economic game against a computer-controlled VH or a visible human opponent. Decisions made throughout the game were logged, additionally participants' faces were filmed during the study and analyzed with expression recognition software. The analysis of choices showed participants are far more willing to violate social norms with VHs: they are more willing to steal and less willing to forgive. Facial expressions show trends that suggest they are treating VHs less socially. The results highlight, that even in impoverished social interactions, VHs have a long way to go before they can evoke truly human-like responses.

Keywords: Virtual humans · Social behavior · Facial expressions · Decision making

1 Introduction

Do people treat machines like people? This has been a central concern within the virtual agent and robotics communities, almost since their inception. The answer to this question has more than passing interest. Virtual humans (VH) are increasingly used to teach people how to interact with other people. VHs teach people how to negotiate [4] or how to overcome fear of public speaking [2]. Others have proposed virtual agents or robots as replacements for people in a variety of customer service and even business settings. Following Cliff Nass' early work on the Media Equation [14], it is common to assume that the same social processes arise in both human and VH interaction, and many subsequent studies have reinforced the validity of this assumption (e.g., [17]).

Yet, recent studies emphasize important differences in how people treat machines [6,9]. Further, there is good reason to believe that studies under-report the differences between humans and artificial partners, as most "direct comparisons" are not as direct as they might seem. The most common method is to manipulate the *mere belief* of who one is interacting with. For example, people

© Springer International Publishing Switzerland 2015
W.-P. Brinkman et al. (Eds.): IVA 2015, LNAI 9238, pp. 452–460, 2015.
DOI: 10.1007/978-3-319-21996-7_48

interact with a digital character but in one case they believe they are playing a computer program and in the other case they believe a person is driving the agents behavior [9]. While there are good methodological reasons to adopt this experimental design, it also clearly under-represents the differences between human and VH interaction. It is a necessary but insufficient step towards demonstrating equivalence between human and machine interaction.

In this study, we make a direct comparison between the behavior of people interacting with other humans in face-to-face interaction with their behavior when interacting with VHs. We explore this within the context of a standard economic game, the iterated prisoner's dilemma, as this allows for several behavioral measures and allows us to connect our findings with a number of existing studies on social behavior. Prior VH research and robotics research on the prisoner's dilemma manipulated the beliefs of participants as to whether they were playing a real or virtual human, but decisions were always made by a computer (e.g., [7,10]). However, in this study we compared data between humans that could see each other via webcam (but not speak to each other), against humans playing with an emotionally-expressive VH. In order to determine social behavior we analyzed both the strategy used by participants and their use of facial expressions, by using facial expression recognition software. Based on prior findings on how people treat VHs in this game [7], we hypothesize that people will be more reluctant to show pro-social behavior to a VH, both through their actions and emotional displays. We explain these hypotheses in the next section.

In Sect. 2 we will give an overview of work involving the displayed social behavior against VHs. Section 3 describes the specifics of the iterated prisoner's dilemma game played during the study, as well as the VH that was used. Furthermore information on the analysis of the game behavior and expressed behavior will be given. Section 4 contains the overview of the results of the study for both game and expressed behavior and in Sect. 5 the implications of these results will be discussed.

2 Related Work

There are several views on the social behavior people show when interacting with computers. The Media Equation of Reeves and Nass claimed that responses to computers would equal responses to humans when computers incorporate human-like social cues. This was claimed to occur because people develop automatic responses to social cues and thus, unconsciously react automatically to computers in the same way as they do towards other humans [13]. It has been argued that the concept of facial expressions within economic games can serve as automatic elicitors of social behavior [16] and that VHs can exploit these cues [5].

A strong interpretation of the media equation, often articulated by Nass [12], is that responses towards computers are equivalent to human responses when computers incorporate social cues. A more nuanced perspective replaces the "=" in Nass' media equation with a "<". Blascovich [3] argues that social influence

will increase based on the perceived realism and "agency" of a virtual agent. Agency refers to the perceived sentience or free will of an agent. This view is supported by a study of de Melo et al. [7], in two experiments agency was manipulated by comparing VHs that were either agents (i.e. controlled by computers) with avatars (i.e. controlled by humans). These experiments showed that people cooperated more with VHs which showed specific facial displays, however these displays only scored significantly different for the avatar condition, thus showing the difference in social behavior people display while playing against humans or VH. Riedl et al. [15] have done a study where participants played against humans and avatars. Their results showed that people display similar trust behavior between humans and avatars, however through neuroimaging they showed that there are different responses in the brain between human and avatar opponents. In a study by Krach et al. [10] humans played an iterated prisoner's dilemma, against both computers, robots and humans. Their results showed that humans in fact experienced more fun and competition in the interaction with increasing human-like features of their partners.

Our current study builds on the findings of de Melo et al. [8]. In their study, people played an iterated prisoners dilemma with a VH that played tit-for-tat and expressed specific emotions. In one condition, participants believed the agents choices and emotions were selected by another participant. In the other, they believed they were generated by a computer programmed to behave like a human. In either case, players could send emotional expressions to the other player along with their choice in the game. The tit-for-tat behavior and the pattern of emotional expressions were both chosen to maximize the amount of cooperation shown by participants. Nonetheless, participants made less cooperative choices and sent fewer positive and more neutral expressions when they believed they were playing a computer opponent. Based on these findings, we make the following hypotheses:

H1: *Participants will cooperate significantly more with other human players than VHs. More specifically, we predict people will be (H1a) more willing to try to exploit a VH, (H1b) more willing to persist in exploiting a VH, and (H1c) more willing to forgive humans following exploitation.*

H2: *Participants will show more cooperative facial expressions to human players compared with VHs. Specifically, we predict people will (H2a) show more joy to human players and (H2b) show more neutral expressions to VH.*

3 Experimental Setup

For this study, participants played an iterated prisoner's dilemma game against either other humans or a VH. The study used a 2-cell design, a total of 113 participants (56 female) participated in this study. 23 participants played against the VH, the remaining 90 participants played the game against each other in the human condition. No specific information was given on the VH, participants

Table 1. Left: Number of tickets the participant receives per outcome. Right: VH responses to outcomes.

		Opponent	
		Cooperate	Defect
Participant	Cooperate	5	0
	Defect	10	1

		Virtual human	
		Cooperate	Defect
Participant	Cooperate	Joy	Fear
	Defect	Anger	Sadness

were simply told that they would play the game against either a human or a virtual human. The task was based on the one presented by de Melo et al. [7]. Participants played 10 rounds of the game and the possible outcomes of the player decisions are shown in Table 1.

The game interface, shown in Fig. 1, displays the game on one side of the screen and the opponent on the other side. The participants chose whether to "split" the tickets or try to "steal" them, corresponding to the cooperate and defect options.

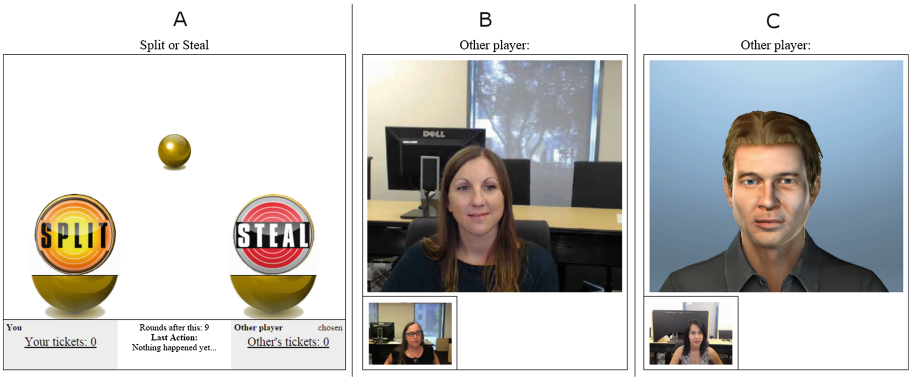

Fig. 1. Screenshot of the split/steal game. Panel A shows the game at the moment the participant can make their choice. Panel B will only be shown to participants in the human condition, panel C only for the VH condition.

Figure 1 shows both the human and VH condition of the game. The experiment was performed in a lab setting, with a maximum of five participants playing the game on computers. Participants were not allowed to speak during the study. The group playing against human opponents could see the video from their opponents' webcam on their screens. Participants playing against the VH would instead see the VH display within the web browser using the Unity web player plugin.[1] The VH used a tit-for-tat strategy during the game, similar to the study by de Melo et al. [7]. The agent used this strategy for the entire game with the exception of the first and second round, on the first round the agent would always cooperate with

[1] https://unity3d.com/webplayer.

the participant, whereas on the second round the VH would always defect. Table 1 shows the facial expressions feedback of the VH on the outcome of a round. These expressions are based on the expressions of virtual agents tested in a study by de Melo et al. [5] and were found to perform the best at eliciting cooperation. The actions of the participants were logged in a database along with the timestamps. Using this data we could infer when decisions were made, when the results were revealed and when rounds began or ended.

Participant videos from the webcams were automatically analyzed using FACET facial expression recognition software.[2] FACET features include intensities for the basic emotion labels as well for overall sentiment labels: "POSITIVE", "NEGATIVE" and "NEUTRAL". FACET is a commercial software for expression recognition that evolved from an academic version, the "Computer Expression Recognition Toolbox" (CERT) [1] and reports high accuracy on emotion recognition labels on known datasets [11]. Videos with high rate of missing frames were automatically discarded from the analysis. Logging of the game events allowed for automatic event-based behavior encoding as well as automatic segregation of the signals on the game period from the overall recording.

4 Results

This section describes the results of our study. Section 4.1 shows our findings on H1, Sect. 4.2 the findings on H2.

4.1 Game Behaviors

The plots in Fig. 2 show how often participants chose to cooperate in both conditions. We performed an independent T-test on this data, which showed

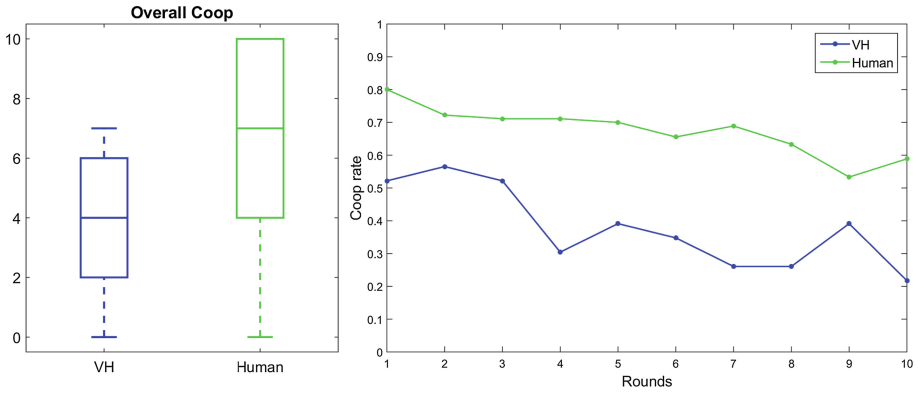

Fig. 2. The left boxplot showing the overall cooperation rate for participants playing either a VH or human opponent, the plot on the right displays the cooperation rate per round

[2] http://www.emotient.com/products/#FACETVision.

that there was a significant difference in overall cooperation between the human ($M = 6.77$, $SD = 3.17$) and the VH conditions ($M = 3.64$, $SD = 2.13$); $t(108) = 4.39$, $p < 0.001$. The number of times participants perform mutual cooperation in the human condition ($M = 5.19$, $SD = 4.07$) is in this game state significantly higher than in the VH condition ($M = 1.59$, $SD = 1.65$); $t(108) = 4.06$, $p < 0.001$, whereas the opposite is true for the mutual defect state (Human: $M = 1.69$, $SD = 2.17$; VH: $M = 4.05$, $SD = 2.44$); $t(108) = 4.44$, $p < 0.001$.

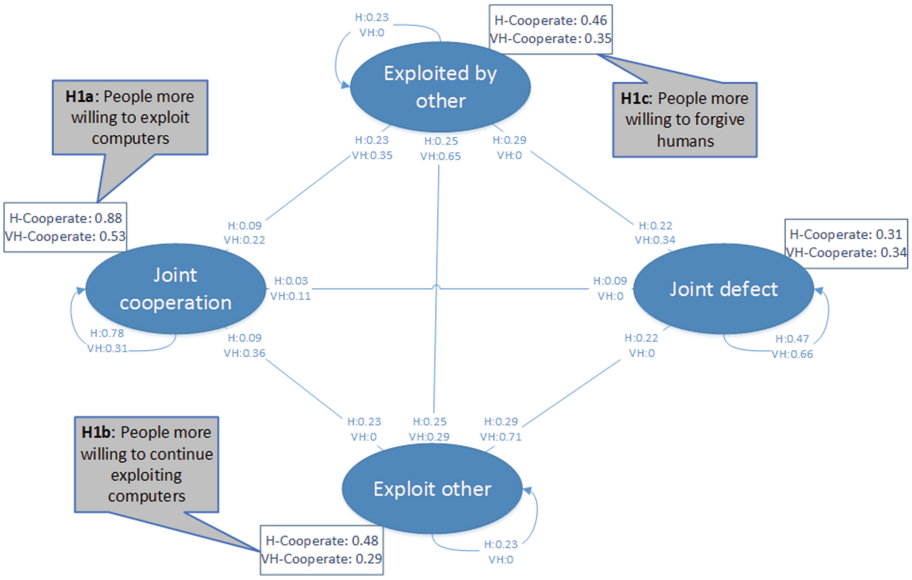

Fig. 3. Markov chain of the possible game states. Boxes display the chance a participant would choose to cooperate given in a certain state and support H1.

The Markov Chain in Fig. 3 shows that participants are generally more likely to exploit VH opponents than real humans (H1a). When participants are in a mutual cooperation state, the probability of continuing to cooperate is 88 % for the human condition, but only 53 % for the VH condition. Participants will forgive human opponents more easily after being exploited (H1c), with a 46 % probability, whereas for VH it is only 35 %. Similarly, participants are more likely to continue defecting on a VH opponent after having already exploited them once (H1b), with a probability of 29 % participants will choose to cooperate again after betraying their VH opponent, while this is 48 % for real humans.

Participants facing other humans are overall more likely to achieve joint cooperation with a probability of 52 % while only 17 % for the VH group. Participants playing against the VH instead had a chance of 39 % to reach mutual defect, whereas for the human group this chance was 17 %.

The self-report questionnaire data also supports the hypothesis. On a 7-point Likert scale people considered themselves significantly more cooperative while

playing against humans (M = 5.88, SD = 1.72) than against VHs (M = 4.57, SD = 1.70), $t(112) = 3.27$, $p = 0.001$. Participants also self-reported that they were more fair against humans (M = 5.54, SD = 1.85; VH: M = 4.30, SD = 1.89), $t(112) = 2.85$, $p = .001$, further supporting H1. We found similar results in the self-reported data on friendliness, honesty and positivity.

4.2 Expressed Behaviors

As a secondary aspect of the behavior towards the game opponent we examined the participant displays of emotion during the game. For this purpose we used the automatically extracted measures mentioned in Sect. 3 and looked mainly at the intensities of the summary labels: "POSITIVE", "NEGATIVE" and "NEUTRAL".

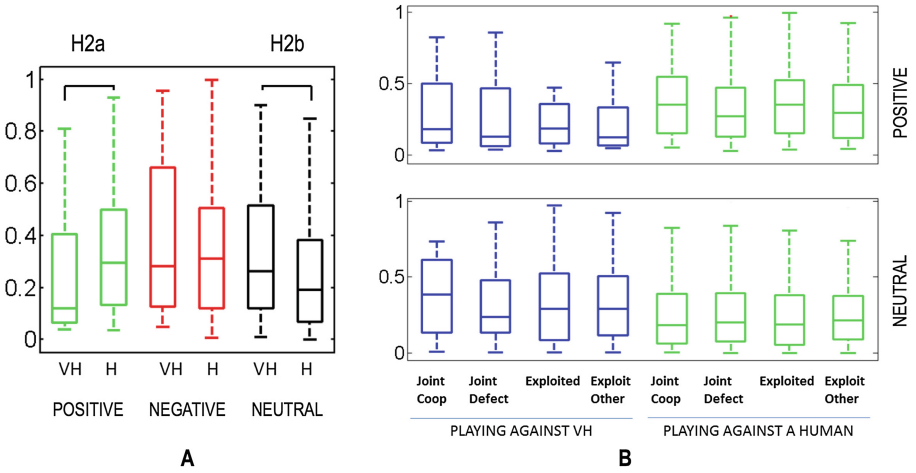

Fig. 4. Comparison of displays of expressions when playing with a VH or a human. Both overall (A) and when breaking down the game by game state (B), participants display more cooperative behaviors on average when playing with a human.

We show our first observations in Fig. 4A, namely that participants display a trend of higher intensity of positivity (H2a) and less neutrality (which translates to more expressivity, H2b) when playing against another human versus playing against a VH, as we hypothesized in H2. Those trends were both significant at the 0.1 level, compared with a standard T-test.

These trends combined translate into an overall more social signal that participants communicate with their expressions to other humans while VH opponents are receiving a less social treatment. This observation still holds when breaking down by different game states as seen in Fig. 4B (bottom) for expressivity and B (top) for positive intensity. These findings agree with de Melo et al. [8] reports on participants' chosen signaled affect during the game and support our second hypothesis that people will display more cooperative expressions to human players compared with VH ones.

5 Discussion

When comparing over the same social dilemma task, we demonstrated that participants will act more cooperatively towards other humans than VHs, both in terms of game choices when choosing to betray, forgive or cooperate (as described in H1) and in terms of displaying cooperative expressions such as more joy, or less neutrality (as described in H2). It can be argued that both of those aspects of behavior form a coherent profile for the players that is more social when facing other human players than when facing VH opponents. This general observation agrees with previous findings [7] that although people treat VH like a social entity, they don't treat them equally to other humans.

The observations made in this study are locally independent of the strategy used by the opponent, however due to the iterative nature of the game the overall strategy used by an opponent should be considered as another confounding factor and be further investigated.

As discussed in Sect. 2 the difference in behavior shown by participants against VHs may have to do with the poorer perception of emotion expression and agency of the VH [3]. Interestingly enough, in the self-report questionnaires the participants reported less connection to a VH opponent ($M = 2.70$, $SD = 1.49$) versus a human opponent ($M = 4.46$, $SD = 1.73$), $t(112) = 4.48$, $p < 0.001$. This less-felt rapport could explain why participants display more neutral expressions while interacting with a VH than with a human opponent. However, considering also the communicative, coordinative role that facial expressions play, one can hypothesize that the knowledge or the expectation that the VH cannot receive these signals the same way as a person does, would lead a person to allocate less effort into that signaling channel and perhaps to cooperative behavior over all. This may be tied with the observation that when playing with the VH, participants scored significantly less than when they were playing with humans (H: $M = 44{:}10$, $SD = 13.24$; VH: $M = 32.83$, $SD = 10.86$), $t(111) = 3.77$, $p < 0.001$. Understanding those gaps better is a topic of future work and it would help bring VH interactions closer to human-to-human ones.

Acknowledgements. This research was supported in part by the AFOSR [FA9550-14-1-0364] and the US Army. The content does not necessarily reflect the position or the policy of any Government, and no official endorsement should be inferred.

References

1. Bartlett, M., Littlewort, G., Wu, T., Movellan, J.: Computer expression recognition toolbox. In: 2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008, pp. 1–2. IEEE (2008)
2. Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., Scherer, S.: Cicero - towards a multimodal virtual audience platform for public speaking training. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 116–128. Springer, Heidelberg (2013)

3. Blascovich, J.: A theoretical model of social influence for increasing the utility of collaborative virtual environments. In: Proceedings of the 4th International Conference on Collaborative Virtual Environments, pp. 25–30. ACM (2002)
4. Broekens, J., Harbers, M., Brinkman, W.-P., Jonker, C.M., Van den Bosch, K., Meyer, J.-J.: Virtual reality negotiation training increases negotiation knowledge and skill. In: Nakano, Y., Neff, M., Paiva, A., Walker, M. (eds.) IVA 2012. LNCS, vol. 7502, pp. 218–230. springer, Heidelberg (2012)
5. de Melo, C.M., Carnevale, P.J., Read, S.J., Gratch, J.: Reading peoples minds from emotion expressions in interdependent decision making. J. Pers. Soc. Psychol. **106**(1), 73–88 (2014)
6. de Melo, C.M., Carnevale, P., Gratch, J.: The influence of emotions in embodied agents on human decision-making. In: Allbeck, J., Badler, N., Bickmore, T., Pelachaud, C., Safonova, A. (eds.) IVA 2010. LNCS, vol. 6356, pp. 357–370. Springer, Heidelberg (2010)
7. de Melo, C.M., Gratch, J., Carnevale, P.J.: The effect of agency on the impact of emotion expressions on people's decision making. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp. 546–551. IEEE (2013)
8. de Melo, C.M., Gratch, J., Carnevale, P.J.: The importance of cognition and affect for artificially intelligent decision makers. In: Twenty-Eighth AAAI Conference on Artificial Intelligence (2014)
9. Fox, J., Ahn, S.J., Janssen, J.H., Yeykelis, L., Segovia, K.Y., Bailenson, J.N.: Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. In: Human Computer Interaction, pp. 1–61 (2014)
10. Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., Kircher, T.: Can machines think? Interaction and perspective taking with robots investigated via FMRI. PLoS ONE **3**(7), e2597 (2008)
11. Littlewort, G., Whitehill, J., Wu, T.-F., Butko, N., Ruvolo, P., Movellan, J., Bartlett, M.: The motion in emotion: a cert based approach to the fera emotion challenge. In: 2011 IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), pp. 897–902. IEEE (2011)
12. Nass, C., Moon, Y.: Machines and mindlessness: social responses to computers. J. Soc. Issues **56**(1), 81–103 (2000)
13. Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 72–78. ACM (1994)
14. Reeves, B., Nass, C.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, New York (1996)
15. Riedl, R., Mohr, P., Kenning, P., Davis, F., Heekeren, H.: Trusting humans and avatars: behavioral and neural evidence (2011)
16. Van Kleef, G.A., De Dreu, C.K.W., Manstead, A.S.R.: An interpersonal approach to emotion in social decision making: the emotions as social information model. Adv. Exp. Soc. Psychol. **42**, 45–96 (2010)
17. von der Pütten, A.M., Krämer, N.C., Gratch, J., Kang, S.-H.: It doesn't matter what you are! explaining social effects of agents and avatars. Comput. Hum. Behav. **26**(6), 1641–1650 (2010)