The illusion of agency : The influence of the agency of an artificial agent on its persuasive power

The Illusion of Agency: The Influence of the Agency of an Artificial Agent on Its Persuasive Power

Cees Midden and Jaap Ham

Human-Technology Interaction, Eindhoven University of Technology,
P.O. Box 513, 5600 MB Eindhoven, The Netherlands
c.j.h.midden@tue.nl

Abstract. Artificial social agents can influence people. However, artificial social agents are not real humans, and people may ascribe less agency to them. Would the persuasive power of a social robot diminish when people ascribe only little agency to it? To investigate this question, we performed an experiment in which participants performed tasks on a washing machine and received feedback from a robot about their energy consumption (e.g., "Your energy consumption is too high"), or factual, non-social feedback. This robot was introduced to participants as (a) an avatar (that was controlled a human in all its feedback actions; high agency), or as (b) an autonomous robot (that controlled its own feedback actions; moderate agency), or as (c) a robot that produced only random feedback; low agency). Results indicated that participants consumed less energy when a robotic social agent gave them feedback than when they received non-social feedback. This behavioral effect was independent of the level of robotic agency. In contrast, a perceived agency measure indicated that the random feedback robot was ascribed the lowest agency rating. These results suggest that the persuasive power of robot behavior is independent of the extent to which the persuadee explicitly ascribes agency to the agent.

Keywords: Persuasive Technology, Agency, Social Robotics, Persuasive Power.

1 Introduction

Imagine the near future in which one of your two personal service robots compliments you on how great you look today. A few minutes later, your other robot approaches you. This second robot has been malfunctioning since the weekend, uttering random sentences. In other words, it clearly is no longer an independent agent that controls what is says. The latter robot has lost some if not most of its agency, that is, it seems to say things involuntarily. You keep reminding yourself that you should hook it up to its maintenance website. While passing, this second robot gives you a compliment on your pleasant demeanor this morning. Now, which of the two robots would receive your warm appreciation in return to its compliment? Would your knowledge about the lowered agency of robot number two decrease the persuasive power of that robot?

M. Bang and E.L. Ragnemalm (Eds.): PERSUASIVE 2012, LNCS 7284, pp. 90–99, 2012.

Earlier research indeed indicates that artificial social agents can have persuasive power. For example, research indicates that when a robotic agent gives participants feedback on their energy consumption in a task in which participants have to make energy consumption decisions, its feedback influences the amount of energy participants consume [5]. More specifically, social feedback from a robotic social agent had stronger persuasive power than comparable feedback coming from a non-social source.

However, artificial social agents are not real humans, and people may ascribe different amounts of agency to them, which can be described as the extent to which a human has the feeling to have interaction with a sentient human being [2, 4]. Although it seems plausible that perceived agency is important to make social behavior of an agent, whether human or non-human, meaningful, recent research has produced some inconsistent results and different views on the effects of agency [see e.g. 9]. In the present paper we will present new research about the effect of agency on the persuasive power of a social agent. First we will briefly review the current state of knowledge regarding social responses to artificial social agents.

Earlier theorizing about the role of embodied virtual agents (EVA's) in social interaction has led to two views (relevant for robot agency effects) that stand out and have been scrutinized empirically in two lines of research. The first view has resulted from the work by Reeves and Nass [15]. Their research suggests that people exhibit social responses when interacting with a computer [e.g., 3, 6, 7, 8, 12, 13]. According to the media equation theory [15], people consider computers as social actors when triggered by certain social cues. These responses are described in the Computer As Social Agents (CASA) paradigm [15]. This paradigm proposes that people react to computers as if they are social actors, while at the same time, people know that these responses are inappropriate (i.e., people know they do not have to react socially to a computer). Numerous studies have supported this notion [for an overview see 15]. For example, it has been shown that people react more politely when a computer asked about its own performance, than when the computer's performance was assessed by a paper-and-pencil questionnaire [13], which is similar to reactions in human-human interaction. However, when asked about their social reactions to the computer, people often deny behaving socially and may even feel offended by the question. In other words, people's implicit reactions (behavior when interacting with computers) and explicit responses when directly asked about computers differ from each other.

Based on such earlier findings, Reeves and Nass hypothesized that humans react in an *automatic* fashion to technical systems if social cues are provided by the system [15]. This explanation is based on a theory by Langer [10] about mindlessness behavior. When in a state of mindlessness, people act on automatic pilot and do not think consciously about their behavior. Attention is directed to information that is relevant for the current task, leaving no room for alternative information [10]. It is suggested that when people interact with computers, they mindlessly form a premature commitment to social scripts of human-human interaction experienced in the past [12]. In short, the CASA paradigm [15] suggests that social cues trigger social rules and people react in accordance with these social rules. However, only few studies tested the effects of social human-agent interactions on persuasion and behavior change. Taken together, the CASA paradigm has demonstrated important psychological effects of human-computer interactions that include social aspects similar to interactions between humans. The hypothesis of automaticity, or mindlessness as suggested by Nass and colleagues [e.g., 12] thereby referring to Langer [10] deserves to be put under empirical scrutiny.

Research investigating the CASA paradigm usually used simple boxed computers, thereby demonstrating the easiness of evoking social responses. This raised the question of how features of the agent that make the agent more realistic, that is, more human-like and more social, moderate the amount of social responses that are triggered in humans interacting with those more human-like artificial agents. This has been the focus of the second type of view on the effects of agency.

Different from the CASA paradigm, this other view has a focus on embodied virtual agents, which usually have humanlike features. For example, the Threshold Model of Social Influence [2] proposes that social human-agent interaction is a function of perceived agency and the social realism of the agent. This model assumes that people react socially to humans or avatars on the basis of social verification, that is, the feeling that they are engaging in a meaningful, social interaction. High agency (perceived for example when interacting with an avatar) makes social realism less relevant. When agency is low, however, social realism may be needed for a human actor to cross the threshold and engage in a social way with an artificial agent. Social realism may be dependent on the behavior and the appearance of the agent [2].

In support of the Threshold Model of Social Influence, earlier research has demonstrated some effects of agency [1] and of social realism [14, 9]. But these results are not consistently found. For example, Vossen, Midden and Ham [17] showed that both speech and appearance of a persuasive agent served as social cues and promoted social influence. However, the two cues did not add up. One cues appeared sufficient to create the effect, suggesting that the cues worked as triggers to start up a social script. Neither is there evidence for the supposed interaction between social realism and agency [see for an overview 9]. Both Nass and Moon [12] and Kraemer [9] indicate that the behavior of the agent is more important than its social features.

In sum, two contrasting conclusions about the relationship between (perceived) agency and the extent to which people engage in a social way in the interaction with an artificial agent can be drawn from these theories. According to the CASA paradigm it does not matter how much agency an agent has, because people will automatically interact with the agent in a social way as long as it shows social cues. However, according to the Threshold Model [2] the effect of an artificial agent will be dependent on its perceived agency and, if perceived agency is not high enough, also on the social realism of the agent. Based on more recent work both models suggest that the richness of social behavioral cues makes a difference on people's social responses. However, differences in research paradigms and design limitations make it difficult to draw clear conclusions about the mechanisms underlying social responses in human-agent interactions. Furthermore, most effects concern psychological effects in terms of perceived presence, liking of the agent and so on. Only very few studies have addressed the potential of artificial agents to influence human behavior through social influence.

1.1 The Current Research

Our experiment was primarily designed to test the effects of agency on behavior change through social feedback. Secondly, we want to compare the effectiveness of feedback by a social realistic agent to the effectiveness of feedback by a non-social system. Thirdly, we wanted to test whether behavioral changes were mediated (or not) by conscious judgments about the attributed social features of the agent .

The research setup was based on earlier research by Midden and Ham [11] in which a robotic agent provided social feedback. In various studies it was demonstrated that social feedback (through social approval and disapproval) was effective in enhancing the saving of energy in household tasks like programming a washing machine. Moreover, by comparing the social feedback to non-social evaluative feedback the behavioral effects could be attributed to the perceived socialness of the agent. Also in the present research a robotic agent (Philips iCat) was used to give social feedback on energy consumption, thereby trying to convince people to use less energy whereas the participants also have the additional goal of producing a clean laundry. The current experiment had four conditions. In the first condition, the robot had the highest level of agency, that is, the robotic agent was functioning as an avatar that was fully controlled by a human actor. In the second condition, the robot had somewhat lower level of agency, that is, the same robotic agent was seemingly acting autonomously. In the third condition, the robot has the lowest level of agency, that is, the robotic agent was controlled (as presented to the user) by a lottery system. Finally, there was a control condition in which the feedback was offered through an energy meter that indicated energy consumption on a simple bar that changed height dependent on energy consumption levels [see also 11]. This latter feedback was also evaluative because the bar could for example reach the top of the scale, or only the middle of the scale, just as the feedback given in all three robot conditions, but was inherently not social in nature. Thereby, social cues were completely absent in the feedback source (the energy meter) used in this condition.

Our research assessed two dependent variables. Firstly, we measured the user's (virtual) energy consumption, which functioned as a measure of persuasion by the feedback source (one of the three types of robot or the energy meter). Secondly, we measured the explicit judgments participants made about their perceived agency and social characteristics of the robotic agent.

This design made it possible to compare different levels of agency, to compare the 'social' conditions with the 'non-social' control group and to test whether the behavioral effects were mediated by conscious judgments or not.

2 Method

2.1 Participants and Design

Eighty-four participants (55 men and 29 women, average age = 21.4 years, SD = 3,5) were randomly assigned to one of four experimental conditions: high agency condition, moderate agency condition, low agency condition, or the non-social feedback condition. The experiments lasted 30 minutes for which participants were paid five Euros. All participants were students at Eindhoven University of Technology.

2.2 Procedure and Materials

Participants were invited to engage in an experiment using a simulated washing machine, and were seated individually in a small room. For participants in robotic feedback conditions, an iCat was positioned on the participants' desk, next a computer. An iCat is a robot developed by the Philips Corporation in the form of the stylized head of a cat that is able, among others, to display social expressions by moving lips, eyes, eyelashes, and eyebrows, and by playing speech files. For participants in the non-social feedback condition, the iCat was not present. For all participants, a simulated washing machine panel was presented in the top half of the screen (see Figure 1). Only for participants in the non-social feedback condition, we added an energy meter to the panel. This energy meter clearly displayed the amount of electricity in kWh corresponding to the chosen washing program. For all participants, in the bottom half of the screen, a program displaying the instructions, tasks and questions was presented.

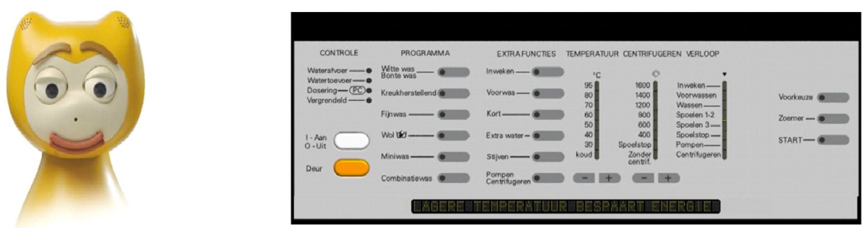

Fig. 1. The iCat and the simulated washing machine interface

This program started with general introductions, and then instructed participants about the task: they were asked to complete several simulated washing trials on a computer. Next, participants were instructed on how to program the washing machine. Participants then were instructed to do each washing as good as possible, that is, to clean the clothes and not damage them (e.g., by washing a woolen sweater at 90°C, causing it to shrink), but also to use as little electricity as possible. So, participants were given two (partly contrasting) goals. To save energy, other settings are optimal (e.g., washing cold) than to clean clothes. Thereby, we motivated participants to use (some) energy, while installing also the motive to save energy for which we could then provide feedback.

To participants in the non-social feedback condition the program next explained the energy meter. To participants in the robotic feedback conditions, the program explained that the iCat would help them save energy by giving feedback on their energy consumption. In high agency condition, a man introduced himself in a video message presented on screen, explaining he would operate the iCat. In the moderate agency condition, the program explained that the iCat was an independent robot, and would give its own assessments in its feedback on the energy consumption of the participant. In the low agency condition, the program explained that chance would determine which feedback would be given, and this was illustrated with an example.

Next, all participants performed a series of washing trials. After one practice trial, the actual 10 washing trials started. For each trial (and also the practice trial), participants were instructed to complete a specific type of wash (e.g., "wash four very dirty jeans"). Each description of a specific type of wash was randomly drawn from a collection of thirty descriptions of common washes, for each trial of each participant such that each participant completed ten different washes. During each washing trial, participants were able to change settings on the washing machine panel until they were satisfied and then pressed a "start" button.

Participants received feedback about the energy consumption of the chosen washing program after each change of settings. Feedback depended on the difference between the chosen settings of the washing machine and a predefined value for that specific washing task. The size of this difference determined the feedback level. We introduced some noise to this difference (by adding a random factor) to the feedback to make the relation between a participant's settings and feedback about their energy consumption less transparent. Thereby we were able to provide also participants in the random robotic feedback condition with correct feedback (at least part of the time) without enhancing perceived agency. In the introduction of the experiment the participants were told how their feedback system worked. For example, they were told the feedback was generated randomly. However, the actual feedback given was identical in all three conditions. Thus, in our setup we kept constant the behaviors of the agents in the three agency conditions.

Participants in the non-social feedback condition received this feedback through the energy meter. The energy meter gave feedback by indicating energy consumption on a scale (on the positive side labeled with "high", "normal", and "low", and on the negative side labeled with "low", "normal", and "high"). Likewise, participants in three social robotic feedback conditions received feedback through the iCat during each trial. For small differences from average energy consumption, the iCat uttered a moderately positive or negative word ("Alright", or "Pretty bad"), for larger deviations a more gravely positive or negative word ("Very good", or "Very bad"), and for large deviations the iCat uttered an even more gravely positive or negative word ("Fantastic", or "Terrible"). All presented labels are best possible translations from Dutch.

After all 10 washing trials had been completed, we assessed the agency judgments of the participants in the robotic feedback conditions ascribed to the iCat. For this, these participants evaluated the iCat on seventeen bipolar 7-point semantic differentials questions. For each of these questions, a word related to low agency was presented on the left of the scale, and a word related to high agency was presented on the right of the scale. Combinations were for example "stupid" and "smart," "non-expert" and "expert," "not social" and "social," and "no intentions" and "full of intentions". A factor analysis of these questions indicated three components with eigenvalues above 1. Only the second (12% explained variance) of these components was clearly related to agency (as it loaded on questions relating to initiative and intentionality). The following items had factor loadings >. 50: passive-active, non-social-social, unconvincing-convincing, untrustworthy-trustworthy, unsupportive-supportive, dead-alive, We used this factor as the agency score in our analyses. The first component (30.2% explained variance) was related quite clearly to questions that tapped perceived competence. The third component (6,6% explained variance) was not clearly related to specific questions. Finally, participants answered demographic questions, were debriefed and thanked for their participation.

3 Results

For each of the ten washing trials of each participant, we calculated the difference between the amount of electricity a participant's settings would have used, and the average usage of electricity for that specific type of wash (e.g., the 4 very dirty jeans) by all participants in our study. We labeled this the energy consumption score. This way, we were able to calculate a dependent variable that indicated the difference between a reference amount of electricity needed for a specific type of washing task (at least in the current study) and the electricity a participant chose to use. To be able to distinguish the effects of positive and negative feedback we calculated an index based on total number of actions of users in the user interface. This means that we not only included the final choices per trial, but all the preceding programming choices. As explained in the method section, these were all followed by (non-social or social) feedback, either positive or negative. The index subtracted for each action the following choice, in terms of energy consumption effect, from the current choice, thereby indicating whether the feedback resulted in a higher or lower energy consumption score for the next following choice.

These scores were submitted to a 4 (feedback condition: high agency feedback vs. moderate agency feedback vs. low agency feedback vs. non-social feedback) x 2 (feedback valence: positive vs. negative) mixed model analysis. This analysis indicated the expected main effect of type of feedback, $F(1, 76.343) = 4.31$, $p < .01$. More specifically, this analysis indicated that non-social feedback resulted in a higher energy consumption score ($M = .09$, $SD = .02$) than in all of the robot feedback conditions, all p's < .03, and presented no evidence that feedback by a robot guided by a human agent (high agency condition) resulted in a different energy consumption score ($M = .04$, $SD = .02$) compared to feedback by an independent robot (moderate agency condition; $M = .01$, $SD = .02$), or feedback by a robot that provided random feedback (low agency condition; $M = .02$, $SD = .02$), all p's > .20.

In line with earlier research suggesting the stronger persuasive effects of negative feedback (a.o. Midden & Ham, 2009), this analysis also indicated that negative feedback resulted in a lower energy consumption score ($M = -.08$, $SD = .01$) than positive feedback ($M = .16$, $SD = .01$), $F(1, 3659.411) = 381.28$, $p < .0001$. This analysis did not indicate an interaction of feedback valence x feedback condition, $F(1, 3654.520) = 1.19$, $p = .31$, indicating that the stronger effect of negative versus positive feedback did not differ for the four types of feedback that we gave participants (non-social, and high, moderate, or low agency feedback).

To analyze effects of the type of agent on participant's agency judgments, we compared participant's agency scores in 1 x 3 (feedback condition: high agency feedback vs. moderate agency feedback vs. low agency feedback) ANOVA. This analysis indicated a main effect of feedback condition, $F(2, 61) = 3.17$, $p < .05$. More specifically, specific contrast analysis indicated that participants' agency judgment scores in the high agency condition ($M = -.28$, $SD = .89$) and in the moderate agency condition were higher ($M = -.27$, $SD = .79$) than participants' agency judgment scores in the low agency condition ($M = -.79$, $SD = .67$), $F(1, 61) = 6.33$, $p = .02$. No evidence was found that participants' agency judgment scores in the high agency condition and in the moderate agency condition differed, $F < 1$. Furthermore, both participants' agency judgment scores for the avatar feedback system differed from participants' agency judgment scores for the random agent feedback, $F(1, 61) = 4.77$, $p < .05$, as did their judgment score for the robot feedback system, $F(1, 61) = 4.66$, $p < .05$. Finally, results indicated no correlation between participants' agency judgment scores and their energy consumption scores, $p > .43$., indicating that agency judgments did not mediate the effects of our agents manipulation on energy consumption behaviors.

4 Conclusion and Discussion

The aim of this research was to investigate whether the persuasive power of an artificial social agent would diminish when the persuadee ascribed less agency to it. To investigate this question, we conducted an experiment in which participants performed tasks on a washing machine and received feedback from a robotic artificial agent about their energy consumption (e.g., "Your energy consumption is too high"), or,non-social feedback. This robot was introduced to participants as (a; high agency condition) an avatar (a human completely controlled all its feedback actions), or as (b; moderate agency condition) an autonomous agent (that controlled all its own feedback actions), or as (c; low agency condition) an agent that (in the eye of the user) provided only random feedback. Results indicated that participants consumed different amounts of energy dependent on the type of feedback. Inspection of the specific contrasts revealed that the greater savings were achieved in the three conditons with social feedback compared to the non-social feedback condition. However no differences in behavioral effects were found between the three agency levels.

Furthermore, results indicated that participants ascribed less agency to the random feedback agent than to the other two agents. However, the difference in perceived agency between avatar and autonomous agent did not reach significance. This finding confirms that, at least at a conscious level, participants were aware that the random feedback agent had less agency than the other two robots. Notwithstanding this awareness of diminished agency, results did not provide evidence that the influence of the social feedback provided by each of the three robots on participants energy consumption choices differed. Furthermore, we could not establish a mediation effect of (explicitly) perceived agency on behavior.

Thereby, the current results provide us with more insight in the cognitive processes of persuasion by artificial agents. Our results indicate that the effect of our three-levels agency manipulation on the behavioral impact of social feedback by the agent did not require an explicit awareness of the level of agency. Furthermore, the agency manipulation did not show a difference at the behavioral level. Apparently, the experience of the agent providing spoken social feedback, as contrasted to the non-social feedback condition, was sufficient to produce the observed behavioral effects. This result seems to support the automaticity hypothesis as proposed in the CASA paradigm [15].

What can be said about the role of social realism as, among others, suggested by the Threshold Model of Social Influence [2]? Our design did not allow for directly testing interaction effects of agency and social realism, but the comparison of the three social agent conditions with the non-social control condition suggests that the higher social realism of the agent's behaviors and appearances compared to the lack of social realism in the control condition (the energy meter gave non-social feedback) contributed to the behavioral effects of the feedback.

In sum, the current research suggests that explicit knowledge indicating higher or lower perceived agency of an artificial agent does not lead to an increase or a decrease of the persuasive power of that agent. The socialness of the agent seems to be more important in cueing the user to activate a social interaction mode. Future research could investigate the additional effects of determinants of social realism of robotic agents. In addition further insight is needed into the different processing modes of assessing agency and social realism by the human user. To return to the description of the two personal service robots you might find present in your home in the near future. The current research suggests that both robots might receive your warm appreciation when they utter a compliment, even when you know that the second robot desperately needs maintenance because it only utters random sentences. Even when an agent seems to have little agency, the social cues that it uses to persuade could be effective, and be able to influence persuadees through automatic, unconscious cognitive processes.

Acknowledgments. We would like to thank Corine Horsch and Bo Merkus for executing the research, and the TUe-HTI Persuasive Technology Lab group for scientific reflection. This research was sponsored by the Dutch scientific research grants organization NL Agency.

References

1. Aharoni, E., Fridlund, A.J.: Social reactions toward people vs. computers: How mere labels shape interactions. Computers in Human Behavior 23, 2175–2189 (2007)
2. 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. Association for Computing Machinery, Bonn (2002)

3. Fogg, B.J., Nass, C.: Silicon sycophants: the effects of computers that flatter. International Journal of Human-Computer Studies 46, 551–561 (1997)
4. Guadagno, R.E., Blascovich, J., Bailenson, J.N., McCall, C.: Virtual humans and persuasion: The Effect of agency and behavioral realism. Media Psychology 10, 1–22 (2007)
5. Ham, J., Midden, C.: A robot that says “Bad!”: Using negative and positive social feedback from a robotic agent to save energy. In: Conference Proceedings of Human-Robot Interaction, 2009, San Diego, USA (2009)
6. Johnson, D., Gardner, J.A.: The media equation and team formation: Further evidence for experience as a moderator. International Journal of Man-Machine Studies 65, 111–124 (2007)
7. Johnson, D., Gardner, J.A.: Exploring mindlessness as an explanation for the media equation: a study of stereotyping in computer tutorials. Personal and Ubiquitous Computing 13, 151–163 (2009)
8. Johnson, D., Gardner, J.A., Wiles, J.: Experience as a moderator of the media equation: the impact of flattery and praise. Int. J. Hum.-Comput. Stud. 61, 237–258 (2004)
9. Krämer, N.C.: Social Effects of Virtual Assistants. A Review of Empirical Results with Regard to Communication. In: Prendinger, H., Lester, J., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 507–508. Springer, Heidelberg (2008)
10. Langer, E.J.: Matters of Mind: Mindfulness/Mindlessness in Perspective. Conscious. and Cogn. 1, 289–305 (1992)
11. Midden, C., Ham, J.: Social influence of a persuasive agent: the role of agent embodiment and evaluative feedback. In: Conference Proceedings of Persuasive 2009, Claremont, USA (2009)
12. Nass, C., Moon, Y.: Machines and Mindlessness: Social Responses to Computers. J. of Soc. Issues 56(1), 81–103 (2000)
13. Nass, C., Moon, Y., Carney, P.: Are respondents polite to computers? Social desirability and direct responses to computers. Journal of Applied Social Psychology 29, 1093–1110 (1999)
14. Nowak, K.L., Biocca, F.: The effect of the agency and anthropomorphism on users’ sense of telepresence, copresence, and social presence in virtual environments. Presence: Teleoperators and Virtual Environments 12, 2–35 (2003)
15. Reeves, B., Nass, C.: The media equation. Cambridge University Press, New York (1996)
16. Roubroeks, M., Ham, J., Midden, C.: When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. International Journal of Social Robotics 3, 155–165 (2011)
17. Vossen, S., Ham, J., Midden, C.: What Makes Social Feedback from a Robot Work? Disentangling the Effect of Speech, Physical Appearance and Evaluation. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) PERSUASIVE 2010. LNCS, vol. 6137, pp. 52–57. Springer, Heidelberg (2010)