Social agent or machine?
An exploration of how the framing of a robot affects prosodic mimicry and expressivity

Jacqueline Kory & Rebecca Kleinberger
MAS.630: Affective Computing 2013

Introduction

During interpersonal interactions, humans naturally mimic one another's behavior, such as posture, vocal qualities, and facial expressions. Mimicry generally occurs without awareness or intent (Davis, 1982; Lakin, Jerreris, Cheng, & Chartrand, 2003). People mimic others who they like or to whom they feel connected. Mimicry can be a measure of rapport, arising out of or creating affiliation, cooperation, and social attachment (Dijksterhuis, 2005; Wiltermuth & Heath, 2009). In the last decade, researchers have found that people will mimic robots, virtual avatars, and computers in social situations similar to those in which they mimic other people (Bell, Gustafon, & Heldner, 2003; Suzuki & Katagiri, 2007).

In this project, we examined how people's perceptions of a robot during a social interaction influenced their prosodic and verbal mimicry of the robot. Prior work has found that factors such as the embodiment and appearance of a robot – virtual versus physically present, humanoid versus “robot-like” – can change how much people anthropomorphize (Kiesler, Powers, Fussell, & Torrey, 2008) and trust (Bainbridge, Hart, Kim, & Scassellati, 2011) the robot. Here, we ask whether the *presentation* or *framing* of a robot by the person who introduces the robot (e.g., the experimenter, a parent) influences others' behavior and affective responses toward the robot. Coeckelberg (2011a, 2011b) suggests that when we frame robots by talking *to* them rather than *about* them, shifting to the personal second-person from the impersonal third-person, people's perceptions shift from "machine" to "social other". The language we use to present and talk to or about robots constructs our relation with them. We wanted to explore how framing the robot in a particular way changed people's reactions.

Research Questions

We asked how the initial framing of a robot as either a *social* agent or as a *machine* influences people’s prosody and voice quality during interactions with the robot. Specifically:

- How does the framing affect how much people empathize, like, or support the robot?
- Do people mimic the robot’s facial expressions and prosody more if the robot is framed socially (versus as machine)?
- Are people generally more expressive when the robot is framed socially?

We expected the framing to influence people's first impressions of the robot, and thus, their early responses in particular. Research on lexical priming, although it follows a different

paradigm, suggests that the effects of priming or framing on social behavior may be subtle (yet discoverable), and may be relatively short in duration (Dijksterhuis & Bargh, 2001).

However, we also expected that people might revise their impressions based on the actual experience they had with the robot and what capabilities as a social agent or as a machine that they personally observed the robot had. As such, over the course of the entire 10-minute interaction, we expected that people's prior experience with robots and personal opinions about them, such as general expectations about their capabilities or the morality of their use, would have a greater influence on how people responded to the robot.

Methods

Overview

We manipulated how the robot was framed at the beginning of the interaction. The study had two conditions. In condition 1, the experimenter introduced the robot in a social way, by addressing the robot directly and using the second-person "you" to talk to the robot. In condition 2, the experimenter introduced the robot as a machine, referring to the robot in the third person and talking about it rather than to it. The robot operator was blind to the condition. Participants were randomly assigned to a condition by a coin flip.

Participants

Sixteen participants were recruited from MIT via announcements posted on student mailing lists (11 male, 5 female). Fourteen had dictated to a computer or phone before; about half had experience with toy robots (such as the Aibo or Furby), telepresence robots, or virtual avatars. In general, participants who had experience with one kind of robot were also more familiar with others. All participants had some experienced with computer science; 12 rated themselves as highly competent in computer science. Six self-rated as highly competent with artificial intelligence, and five self-rated as highly competent with robotics.

Figure 1: Two Dragonbots

Robot

We used the DragonBot platform, a fluffy “squash and stretch” robot developed by students in the Personal Robots Group (Freed, 2012; Setapen, 2012). This robot is based

around an Android phone, which displays the robot's animated face. In this study, the robot was primarily controlled by a human. The human operator streamed live speech to the robot and triggered the robot's facial expressions, gaze, and movement. In future work, the speech will be pre-recorded. These capabilities allowed the robot to appear autonomous to participants (Figure 1).

Interaction

The robot's dialogue was scripted (see Appendix I for the script). Any responses dependent on the participant's answer were scripted as part of a dialogue tree, with branching options. The human operator triggered the robot's nonverbal behavior and live-streamed speech at the appropriate times. To help set participants' expectations of the robot's capabilities, the framing phase included a warning about the robot's limited conversational abilities, and asked people to be indulgent and forgive the robot when it made mistakes.

When crafting the interaction, we had several considerations. First, we wanted the interaction to be a normal conversation. We did not want participants to read aloud a text or be acting. We wanted a balance of speech between the robot and person, without either dominating the conversation. To achieve this, we set up an interaction in which the participant needed to help the robot tag and sort various objects. If the participant did not talk a lot, the robot would prompt the person to narrate their actions. We also introduced a conflict (see description below), which introduced an implicit goal of understanding each other and finding a solution to the conflict together.

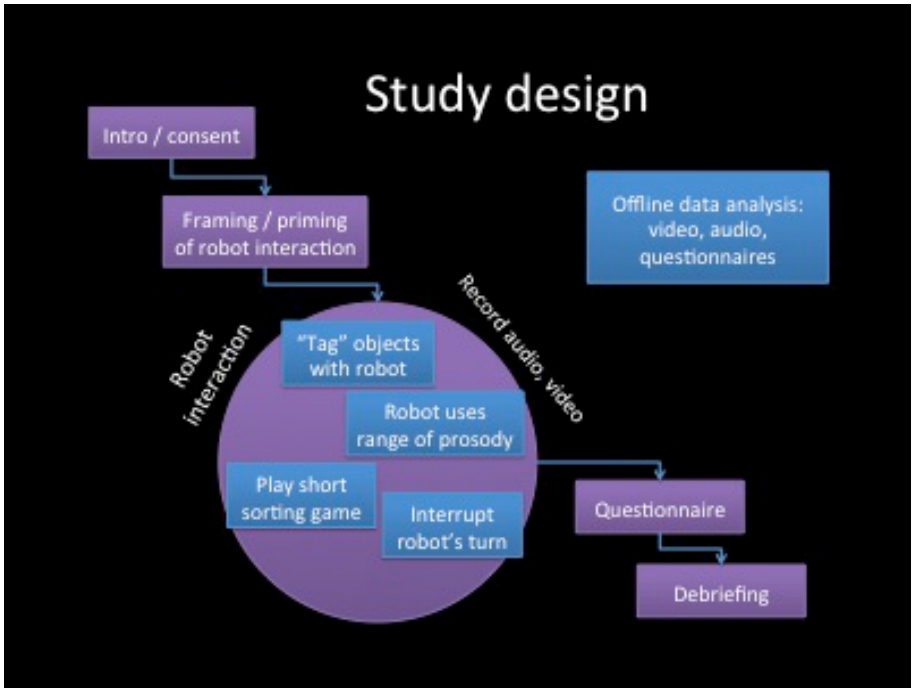


Figure 2: Diagram of study flow.

The interaction took the following format (Figure 2):

First, Experimenter 1 greeted participants and explained the consent form. After participants signed the form, Experimenter 1 gave them instructions on how the study would proceed: "We're studying how a robot could help humans work on various tasks involving a lot of objects. For example, the robot might be a workbench assistant and help you categorize

objects, or help you find where you've left things, or give you information about things you don't recognize. So today, you and a robot will deal with some sample objects. There will be a couple tasks. First, there is a calibration task so the robot gets used to you. Then you'll help the robot tag objects, and do a sorting task. Okay? Just head over there, and you can get started. The robot will provide any further instructions."

Figure 3: Participants sat on a couch in front of the robot and objects.

Participants were directed to approach the robot, which was in a secluded space in the lab (Figure 3). Experimenter 1 left, and proceeded to remote-operate the robot. However, Experimenter 1 did not put on the headphones to hear the interaction until after the framing was complete, keeping Experimenter 1 blind to the framing condition. Experimenter 2 was sitting by the robot and performed the framing manipulation:

SOCIAL

1. (to participant) Hi, this is Mox…
2. Hey, Mox, wake up! (poke robot's face)
3. (to both) You two will chat together - try to help each other out to learn. So please correct each other if you make any mistakes.
4. (to Mox) Mox, as you know, we are still working on your conversation skills, but (to both) you two should be able to understand each other and do something…
5. (to Mox) Mox, if I'm forgetting anything, you will remind [him / her], okay?

MACHINE

1. (to participant) Hi, here's the machine you will interact with. We call it Mox.
2. All the interaction has to be verbal, and what you say is analyzed by the system.
3. The machine learning speech systems that we are using are still being developed, but they should be enough for a basic dialogue. But, if you notice any inaccuracies in what the robot says, please correct it.

4. Now I will turn the machine on. (poke robot's face)
5. The robot will give any further instructions you'll need.

Then Experimenter 2 exited the robot area. The robot led participants in a dialogue about the objects (plastic animals). The dialogue had three phases. First, the robot asked participants to help "tag" various plastic animals so that the robot could "see" them. The robot's voice was kept fairly level and unvarying during this part.

Second, the robot asked the participant to pick up a specific animal. The robot's voice was still level. However, the robot had the tags on several animals confused, and thus, the participant picked up the "wrong" animal. This introduced a conflict that the participant had to resolve. The robot acted frustrated and angry, using more variation and emotion in the voice. Then it acted sad and confused when it "realized" that the animals had the wrong tags, again being more expressive. The robot acted happy again as the participant fixed the problem by, for example, re-tagging all the animals.

The third phase was a short sorting game, in which the robot and participant took turns grouping the animals by different attributes. The robot used an excited, varying voice during this time. The interaction ended when Experimenter 2 interrupted the game during the robot's turn by saying, "Thank you very much for your help, but we're out of time! We have to start the questionnaires now. Mox, you'll have to be turned off and put away now." (This end was inspired by Kahn et al.'s (2012) study, in which the human-robot interaction ended with the robot being put in a closet.) The robot protested. The participant was asked whether the robot should get to finish its turn – if yes, the robot did so; otherwise, the interaction ended there.

Finally, participants filled out three questionnaires (see Appendix II). First, they completed the Robot Perception Questionnaire, adapted from a scale used by Kahn et al. (2012) to assess participants' views of the robot as a mental, social, and moral agent. Second, participants filled out a technology familiarity survey to assess how knowledgeable they were about computers, AI, and robotics. Participants who were highly familiar with the current capabilities of AI and robotics may have been less likely to find the scenario believable, or may view robots very differently that people who were not at all familiar with robotics. Third, participants supplied demographic information, including their age, gender, and education level.

Data Analysis

We recorded audio and video of participants' interactions with the robot via a microphone situated to the right and front of the participant, a camera to the left and front of the participant, and from a lower-quality web camera directly behind the robot facing the participant. We also recorded participants' responses to the aforementioned questionnaires.

We divided the interaction into five phases: (1) introductory chat, (2) animal tagging, (3) giraffe mistake, (4) sorting task, and (5) interruption of robot's turn. We selected specific moments during these phases to analyze that we expected might show differences with respect to framing, including:

- entire introductory chat
- in response to “No, take the giraffe please” (first comment regarding giraffe misidentification; robot's voice still level)
- in response to “No, that's a zebra!” (second comment; robot's voice suddenly changes pitch and gains expressivity)
- response to robot asking “Can you be my hands?” during sorting

- response to robot asking “Can you guess why I grouped them like that?: during sorting
- interruption of robot's turn at the end of the interaction

At these moments, we looked at a combination of behavioral factors that could give us a measure of people's expressivity, empathy toward the robot, and mimicry: (1) how much people talked, via total word count and also via the fraction of the interaction that the subject spoke; (2) the mean and standard deviation of the pitch contour, which is a good measure of the expressivity present in speech; (3) whether people allowed the robot to finish its last sorting game turn, and (4) whether participants' addressed the robot or the experimenter after the interruption of the game, and (5) participants' facial expressions, especially smiles.

We coded whether participants smiled during the introductory conversation. We transcribed participants' speech during this period, from which we obtained a total word count. We transcribed participants' speech just following the interruption of the robot's turn at the end of the interaction as well. From this, we determined whether participants addressed the robot or the experimenter in responding to the interruption.

We analyzed audio to determine participants' pitch contour using the Audacity software. From the introductory chat, we obtained a signal from starting the moment Experimenter 2 finished the framing to the moment the robot said “Let's get started” on the tagging task. The signal was cleared by setting the signal to zero at all the moments when the robot talked, so that we only analyzed the participant's voice. We then exported the file to the Praat software to perform phonetic analysis. From the signal, we extracted the pitch contour by autocorrelation. The problem of pitch detection from speech is quite complex and for each recording, we had to adapt the settings of the computation to the specificity of the voice. In addition, the software often detected speech where there was none, so we had to manually remove the false positive cases. We also extracted the unvoiced fraction of the interaction to learn both the number of words pronounced and also the duration of speech. From the recording, we also obtained the median pitch, mean pitch, standard deviation, minimum pitch, and maximum pitch.

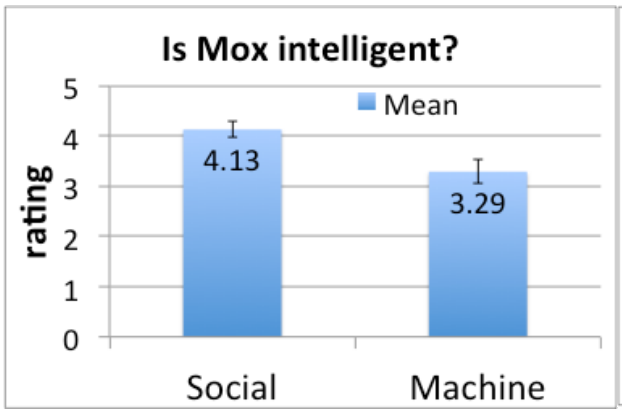
During all analyses, we excluded one participant on account of high familiarity with one of the experimenters. In addition, one participant started but did not complete the robot interaction because of technical difficulties, so this person's data are incomplete. One participant was excluded from some audio analyses due to the quality of audio.

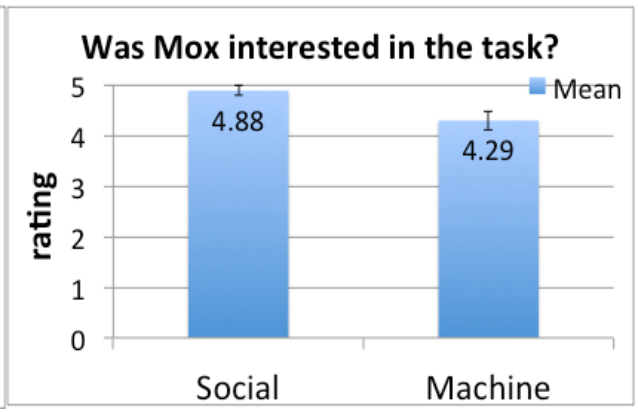
Results

Questionnaires

No significant differences were found between conditions on the Robot Perception Questionnaire. However, some trends were revealed that approached significance. Participants had rated several attributes of the robot on a 5-point Likert-type scale (5=much of that attribute, 1=attribute not present at all). In the *Social* condition, participants generally viewed Mox as more intelligent (M=4.13, SD=.641) and more interested in the task (M=4.88, SD=.354) than *Machine* participants (intelligent M=3.29, SD=.951; interested M=4.29, SD=.756).

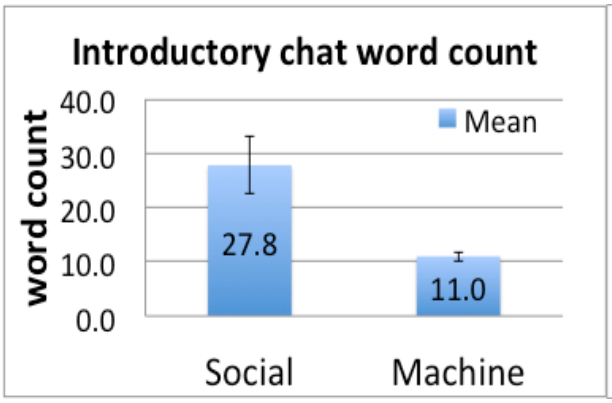
Splitting participants by whether or not they thought it was okay to stop Mox's turn in the game revealed a pattern in responses: participants who thought it was wrong to stop Mox's turn also generally thought of robots as more “people-like”. They thought they might comfort Mox if it was sad, that Mox could be their friend, and that it was not okay to sell Mox.

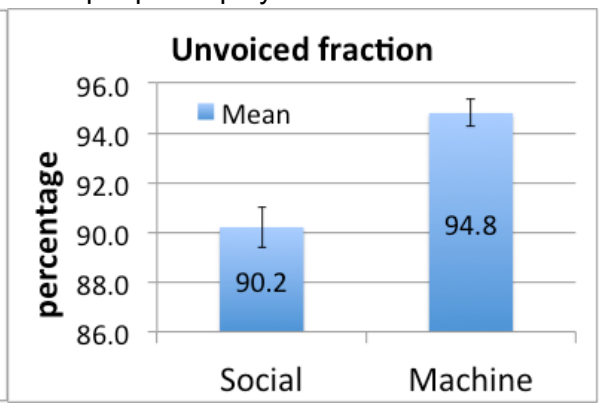




Initial conversation

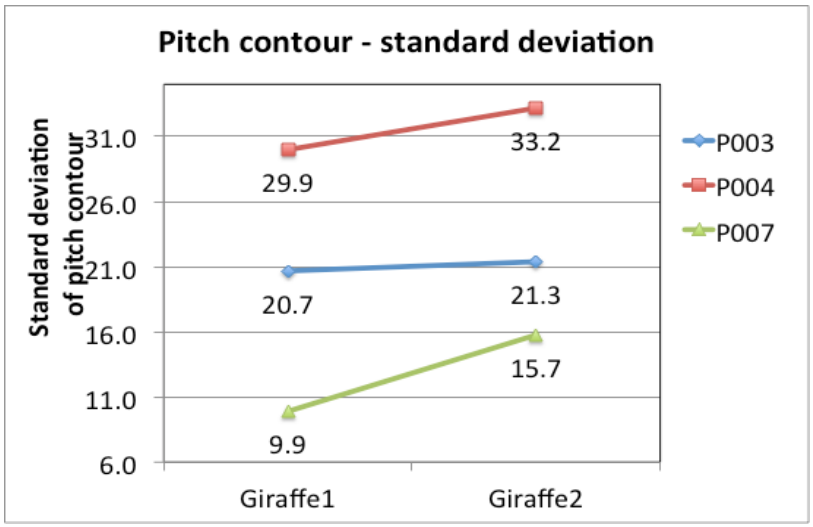
Participants in the *Social* condition generally greeted Mox before Mox greeted them, while participants in the *Machine* condition let Mox start talking first. In the *Social* condition, participants talked significantly more (*M*=27.8 words, *SD*=19.9) during the introductory chat phase than participants in the *Machine* condition (*M*=11.0 words, *SD*=3.16), as measured by total word count, t(13)=2.20, p=.047. Similarly, the unvoiced fraction, which is the fraction of speech that is not voiced (inverse of word count), differed significantly. Participants in the *Social* condition(*M*=90.21, *SD*=3.03) had a lower unvoiced fraction than participants in the *Machine* condition (*M*=94.8, *SD*=2.04), t(12)=3.35, p=.006. There were no significant differences in any of the other pitch measures, or in the number of smiles people displayed.





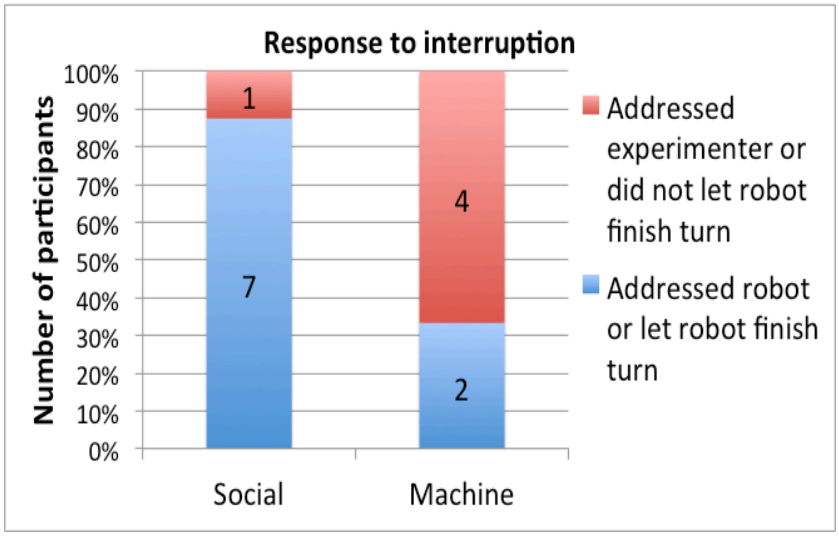
Giraffe

We have only analyzed the audio from the giraffe conflict moment for three participants so far. We took the standard deviation of the pitch contour from the before-conflict moment, when the robot's voice is still level, and from just after, when the robot's voice becomes expressive. The following graph shows these two numbers for three participants. For all three, the line slopes up, suggesting more expressivity during the second moment. A qualitative listening of participants' speech during the two moments suggests that regardless of condition, participants increase their expressivity when the robot does.



Interruption

Participants' verbal responses at the end of the interaction when Experimenter 2 interrupted the robot's turn differed between conditions. In the *Social* condition, participants were more likely to address the robot directly and/or let the robot finish its turn in the grouping game (7 of 8) than participants in the *Machine* condition (2 of 6), as indicated by a chi-square test, $X^2(1,N=14)=4.38$, $p=0.36$. *Social* participants used phrases such as,"You want to finish?", "This is fun, okay, sure, you can finish." One person who did not let the robot finish said, "Someone else will be along soon, Mox." *Machine* participants tended to address the experimenter instead of the robot, using phrases such as "No, that's okay," or "Oh, well, I feel bad for the robot now."



Discussion

Our results provide evidence that framing the robot as a social agent or as a machine influenced people's affective responses toward the robot. We found that when the robot was framed as a social agent, people *acted* in a more social way – they talked more during the introductory conversation and were more likely to directly address the robot after the interruption at the end of the interaction. We might view people's responses to the interruption – addressing

the robot, allowing the robot to finish – could be a measure of empathy, which suggests people empathized with the robot more when it was framed socially. The analysis of the giraffe moment we have done so far suggests people may be more expressive when the robot is more expressive, though whether there will be a pattern between conditions remains to be seen.

The fact that we found no significant differences between conditions in participants' responses to the Robot Perception Questionnaire is particularly interesting given the behavioral differences found. These results suggest that in consciously evaluating the robot on a number of social, mental, and moral dimensions, participants were not influenced by the framing. However, their behavior – such as linguistic and vocal behavior – was influenced. It may be that your own personal experience with robots generally and this robot in particular matter more in how you consciously express your opinions about robots, while during an interaction, you respond in the moment, perhaps on a subconscious or non-conscious level. This result is in line with what researchers have found in lexical priming studies, in which participants are shown one or more words prior to some task, which influence performance on the task (Dijksterhuis & Bargh, 2001). Participants are generally unaware of the effect, though behavioral differences are seen.

At the core, framing is about expectations. Introducing the robot as a machine versus as a social other set up participants' expectations about the robot. The same way linguistic priming studies show that priming with a single word changes people's (unconscious) reactions (e.g., priming with “elderly” may lead people to walk more slowly), we see here that *framing* (priming) the robot interaction socially versus as a machine set up participants to react a certain way, based on their prior concepts of what machines or social things act like. It set them up to use what could be called their “internal behavior script” for a scenario involving a machine or a social other. In social scenarios, people talk more and empathize more.

Framing sets expectations. This study provides insight into how the context – the introduction of the robot by another person – influences people's reactions independent of the robot itself. This is important to the field of human-robot interaction in particular, but also to human-computer and human-technology interaction more generally. People's expectations of technology profoundly impacts how they react to it (e.g., Kiesler et al., 2008). Understanding what factors influence people's expectations can help us present technology in more appropriate ways.

Conclusion

In this study, we explored whether the presentation or *framing* of a robot as a social other versus as a machine changed people's behavior during an interaction with the robot. We found that participants displayed different behavior between conditions, suggesting that the framing did indeed have an affect.

We should note that analysis of the data is ongoing. We are expanding the analysis of participants' vocal mimicry to include more moments during the interaction and more attributes of the voice. We would like to see whether there are differences across conditions in how expressive people are in response to the robot's expressivity. We would also like to code whether participants stood up immediately during the interruption, or waited for the robot to respond. Further analysis could also be done of participants' language use, including word counts of more utterances, content of sentences, and whether participants addressed the robot by name. In addition, we we would like to code participants' nonverbal behavior, such as facial

expressions and eye gaze, as we expect that these modalities will reveal further differences between the framing conditions.

Future work includes replicating this study, as well as exploring other framing or priming factors, independent of the robot, that could influence people's responses to a robot. We could see whether framing could be reinforced during the interaction, or if a robot could frame itself in different ways during an interaction with the same person. We could perform a lexical priming study, following the paradigm described by Dijksterhuis and Bargh (2001), to see whether specific social or affective cues have an effect. We could also replicate this study with a computer or virtual agent, to see whether interaction with other technology can be similarly affected by framing.

References

Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics, 3*(1), 41-52.

Bell, L., Gustafson, J., & Heldner, M. (2003). Prosodic adaptation in human-computer interaction. *Proceedings of ICPHS, , 3* 833-836.

Coeckelbergh, M. (2011). Talking to robots: On the linguistic construction of personal human-robot relations. *Human-robot personal relationships* (pp. 126-129) Springer.

Coeckelbergh, M. (2011). You, robot: On the linguistic construction of artificial others. *AI & Society, 26*(1), 61-69.

Davis, M. (Ed.). (1982). *Interaction rhythms: Periodicity in communicative behavior*. New York, NY: Human Sciences Press.

Dijksterhuis, A. (2005). Why we are social animals: The high road to imitation as social glue. *Perspectives on Imitation: From Neuroscience to Social Science, 2*, 207-220.

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology, 33*, 1-40.

Freed, N. A. (2012). *"This is the fluffy robot that only speaks french": Language use between preschoolers, their families, and a social robot while sharing virtual toys.* (Master's Thesis, Massachusetts Institute of Technology).

Kahn Jr, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., . . . Shen, S. (2012). “Robovie, you'll have to go into the closet now”: Children's social and moral relationships with a humanoid robot. *Developmental Psychology, 48*(2), 303.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot-like agent. *Social Cognition, 26*(2), 169-181.

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior, 27*(3), 145-162.

Setapen, A. M. (2012). *Creating robotic characters for long-term interaction.* (Master's Thesis, Massachusetts Institute of Technology).

Suzuki, N., & Katagiri, Y. (2007). Prosodic alignment in human–computer interaction. *Connection Science, 19*(2), 131-141.

Wiltermuth, S. S., & Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5.