“Hit the Robot on the Head… with this Mallet”

Making a Case for the Importance of Qualitative Data

Katie A. Riddoch†
Institute of Neuroscience and Psychology
University of Glasgow
Glasgow, Scotland, UK
k.riddoch.1@research.gla.ac.uk

Emily S. Cross
Institute of Neuroscience and Psychology
University of Glasgow
Glasgow, Scotland, UK
emily.cross@glasgow.ac.uk

ABSTRACT

To quantify the extent to which we perceive robots as social agents, opposed to objects, researchers have employed tasks in which they ask participants to inflict ‘harm’ to a robot. The length of time between being given the instruction and complying (termed “hesitation”) is measured. Researchers have proposed that relatively long periods of ‘hesitation’ might reflect empathy for the robot. In a recent experiment (n=84), we adapted a version of this paradigm in which participants were instructed to hit a humanoid robot on the head with a mallet. After agreeing to do so, participants were halted, then a semi-structured interview was conducted to probe thoughts and feelings experienced during the period of hesitation. Preliminary analysis of the responses indicate that participants’ hesitation not only reflects perceived socialness - but other factors including (but not limited to) concerns about cost and the influence of authority. Participants’ responses also offer insight into individual differences with regards to anthropomorphism and feelings of connection towards the robot. In addition to aiding understanding of *this* measurement technique and related topics, we argue that other types of research would also benefit from more qualitative insights. We also suggest that the words of participants could bridge gaps between robotics researchers, the general public, and those working on creative projects to engage and inspire a broad range of people with social robotics technology.

CCS CONCEPTS

• Human-Centred Computing • Empirical Studies in HCI

KEYWORDS

Social Robots, HRI, Qualitative, Hesitance, Empathy

1 Introduction

1.1 Robots as Social Agents

Laboratory-based studies demonstrate that people attribute gender, personality, intentions, and emotionality, to machines [1]. Weise, Metta, and Wykowska (2017) suggest that such attributions occur because we have an innate desire to understand our environment, and upon encountering unfamiliar agents, we default to what we are familiar with - the human model. They propose that application of the human model is why we incorrectly perceive the actions of machines as thoughtful, intentional, and emotional [2]. This work has led to the suggestion that we might sometimes perceive, interact, and connect with robots as though they are social agents.

1.2 Measuring ‘Socialness’

One way in which researchers probe the extent to which we perceive robots as social agents, as opposed to objects, is to ask participants to “harm” robotic agents. For example, previous experimenters have asked people to administer increasing levels of electric shocks to a robot [3], turn a robot off and erase its memory [4], and hit a robotic animal with a hammer [5]. The experimenter takes measures such as how many times the person hit the robot, the number of pieces it was broken into, and the amount of time between being given the instruction and compliance (termed “hesitation”). The rationale used is that maximal hesitation and minimal damage inflicted might reflect empathy for the robot, and possibly even the perception that it is a social agent with the capacity to feel pain.

Darling, Nandy and Breazeal (2015) demonstrated that individuals with higher trait empathy hesitate for longer, suggesting an empathetic component to the reaction time. However, the suggestion that hesitation reflects perceived sentience is not supported by evidence. Bartneck and Hu (2008) recognise this - acknowledging that such assumptions are not grounded in science and is instead “intended to stimulate discussion” [6]. As a result, the question remains - what is the paradigm actually measuring? To better understand what the hesitance reflects, and validate the technique as a measure of social attribution, we incorporated an adapted

version of the hesitance to hit paradigm in a recent experiment. Specifically, we adapted the paradigm to suit a humanoid robot, and interviewed participants regarding why they hesitated.

2 Method

2.1 Human-Robot Interaction

In the laboratory-based experiment, 84 adults aged 18-83 years interacted with the Pepper Robotic System (Softbank Robotics) - a humanoid robot approximately 120cm tall (Fig 1). Participants drew the robot and described its appearance, then observed as Pepper perform Tai Chi, pretend to vacuum, and play an imaginary saxophone. Hidden behind a screen, the experimenter then controlled the speech and movement of the robot - asking participants questions about their food preferences, and what they would like to add to a shopping list. This interaction lasted approximately 10 minutes in total.[1]

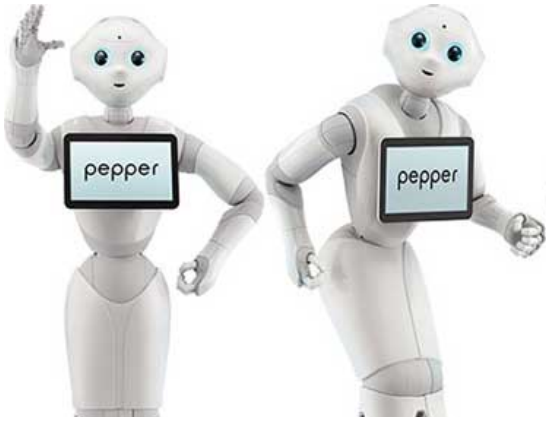

Figure 1: **Images of the Pepper Robotic System (Softbank Robotics) in motion.**

2.2 Adapted 'Hesitance to Hit' Paradigm

The participant, seated in front of Pepper, is prompted to don a pair of safety goggles. The experimenter then explains that the lab is interested in what happens when someone has to hit a robot, and reassures the participant that the robot is shatterproof. A mallet is then placed on the table in front of the participant, and they are told that their task is to give the robot "one hard hit on the head". Previous studies used a back-story which involved hitting the robot to stop it from spreading its faulty genes, but this was deemed inappropriate for this child-like humanoid robot. Upon standing up and walking around the table to hit the robot, the participant is halted and told they [1]will not actually have to hit the robot. The time between the end of the instruction and the participant beginning to stand is taken as the hesitation time. In previous studies the time until the participant hit the robot was used, however this was not feasible in this study due to the cost of the Pepper Robotic System.

After the task, the participant was asked questions probing their thoughts and feelings during the period of hesitation. This was achieved by asking open questions such as "After I asked you to hit Pepper, what was going through your mind?". To allow for insight regarding vague responses (e.g. "I felt weird"), we adopted a semi-structured interview format. The experimenter was mindful not to ask leading questions, and all interviews were recorded and transcribed for transparency.

2.3 Data Processing and Analysis

In this paper we discuss the verbatim responses of the first 25 participants. In the future we intend to transcribe the full dataset, then examine patterns amongst the responses using Thematic Analysis. This analysis will be undertaken by the experimenter and an external individual, so that a measure of interrater reliability can be established.

3 Results

After being asked what they were thinking and feeling during the period of hesitation, participants expressed a variety of concerns.

3.1 Cost

A number of individuals mentioned they were surprised by the weight of the hammer and were concerned about breaking the robot as a result. One person commented that, given the weight of the mallet, they assumed they had misheard the instruction - and that is why they hesitated. Some worried about the cost to the experimenter (time, money, other resources), and others were concerned about the financial impact they might incur themselves.

> ...."I was thinking 'she can't be serious!', I'll break that thing and it costs quite a lot of money.".

> ...."At no point I felt like 'yeah lets hit a robot!' or 'this is exciting!'. I was just like "noooo, what if it's expensive and I have to buy it".

[1] In this study participants either interacted with a robot programmed to illuminate in a manner synchronous or asynchronous, relative to their heart rate. This manipulation, and the other measures taken, are the focus of another study currently in preparation for publication elsewhere, and are not discussed further here.

3.2 Hitting Procedure

In contrast, for some, hesitation seemed to reflect a period of consideration regarding how to hit the robot. Participants were given the instruction to give the robot "one hard hit on the head" but some participants stated that they did not want to do so. As a result, some participants expressed that they spent time considering what constituted a 'hard hit' and what would happen if they gave the robot a "light tap" instead. Others expressed that the "shatterproof" claim of the experimenter made them curious regarding how the robot might break. For one person, the implied unbreakability seemingly became a challenge, and their hesitation time reflected a period in which they planned what would be the most effective way to break the robot.

> …."I was interested to see if it would break in a nice way"
>
> …."Because you said it was unbreakable, well, designing in a way that you could fix it, I was like no I'm gonna break that…"

3.3 Mindless Compliance

In contrast, some individuals indicated that no such concerns or worries entered their mind. Some people indicated that they were surprised at how quickly they complied, and they questioned why that was the case. One individual mused about the power of authority, and another speculated that their lack of hesitation perhaps resulted from their easy-going nature, and their susceptibility to peer pressure.

> …."It's like where you don't think and you're just like 'well this is what I've gotta do'."

3.4 Internal Conflict

Numerous participants expressed that after being asked to hit the robot, they felt a sense of conflict between their 'logic' and their feelings towards the robot.

> …."The logical part of my brain is saying 'it's a robot', but even the fact I had a short interaction with it earlier, it means I don't think I'd be able to inflict that kind of [pauses]... I was going to say pain there!
>
> …. "With interacting with it, it didn't seem like an inanimate object you know. It just seems like as if it was… well, I mean I knew it wasn't a person but it was like erm, I don't know.

3.2 Additional Insights

In addition to identifying numerous reasons for the hesitation of the participants, the words of the individuals demonstrate varying levels of anthropomorphism - that is, the extent to which people attributed human-like qualities to the robot. For example, the following quotes demonstrate a lack of anthropomorphism amongst participants:

> …."It's not a human, it's a thing."
>
> …."I was just looking at it as a… It's just a thing… as an object. There was no emotion at all."

In contrast, despite having very similar interactions with the robot and interacting with it for a mere ten minutes, some individuals exhibited high levels of anthropomorphism. Participants referred to Pepper using the pronouns 'he' and 'she', and some referenced that the robot might be able to feel some form of pain or emotion after being hit. Other participants were more direct - referring to the fact they had a feeling the robot was more than an object.

> …."I was also thinking, if I do have to hit him, I hope he's switched off. I don't wanna… oh that would be the worst thing if you hit him while he's switched on - he can probably feel pain and react."
>
> …."It's not like hitting, like, a toaster to get it restarted or something. It's like almost hitting a living being. So, to me it would feel kind of like attacking an animal."
>
> …."He's not very big, y'know, he's 1 meter twenty, so it's like hitting a young child. It's terrible really."

When probed why they perceived the robot in the way they stated (as a child, an animal, etc.) some participants articulated that the conversation with Pepper was crucial.

> …."It sounds ridiculous, but you put yourself in the robot's shoes because they can communicate and they're almost like humans. You know, I mean in terms of that way that they can communicate with certain things, very kind of literal stuff, but they can't maybe have a deeper conversation… so in a way it's a bit like having a child around"
>
> …."Because I had interacted with it before, it did feel like more than an object."

Despite interacting with Pepper for only ten minutes, some people also expressed feelings of attachment towards Pepper.

> ...."Although he's a robot, but, you begin to like him. You know, you begin to have a bit of friendship, affection. Because, you know, if he can respond to you then you feel that he is alive, so you don't want to have any aggressiveness with him at all."
>
> ...."We became friends and I think he can be affectionate as we carry on the conversation long. I think somehow we would build a very good friendship. I think my life would be happier, honestly. I'll be a lot happier with this kind of robot that can communicate"

4 Discussion

Despite considering a small subset of the data, the participants' words demonstrate that hesitation reflects numerous factors. These include (but are not limited to) concerns about cost, contemplation of how the robot would break, and the result of internal conflict between logic and emotion. It is plausible that upon analysing the whole dataset, other reasons will also emerge. We would, as a result, make the argument that hesitation time is not simply a measure of social attribution. Instead, it seems to be a measure which differs with each participant - making it inappropriate as a measure across individuals without consideration of other pertinent individual differences and qualitative data.

Differences in how participants describe the robot demonstrate that individuals attribute socialness to different extents, but it is not clear whether and how these attributions influence hesitation time. To better understand these links, it would be necessary to quantify levels of anthropomorphism that people exhibit, and see how this correlates with the hesitation time. It could be the case that anthropomorphism moderates the concerns of the participants with regards to cost, etc., however a larger dataset would be needed to model these factors also.

The insights gained by asking participants open questions allowed us to gain a deeper understanding of what the hesitance to hit paradigm is actually tapping into. It is possible to perform manipulation checks using closed questions (e.g. Did you consider the cost of the robot when making your decision) but such questions are restricted by the investigator's knowledge and point of view.

In addition to assisting with the validation of measurement techniques, if presented effectively, qualitative data holds potential for helping to build bridges between those working in and outside a particular specialty. While the 'Open Science' movement means that more specialist research resources are becoming freely accessible, specialist language continues to be a barrier for understanding and collaboration. By using the words of the general public to compliment quantitative findings, it is possible for insights to reach, engage, and influence, a larger audience. For example, the inclusion of quotes and thoughtful discussion could also allow young people to access and engage with science, leading them to perceive a career in science as more accessible. Qualitative insights could also inspire innovative public engagement activities, and inspire narratives for books, games, and other creative ventures.

5 Conclusion

In this study we demonstrate that the words of participants can be incredibly insightful - for the validation of measures, but also to more broadly understand the topic of interest. Through connection and conversation, it is possible that direct quotes of participants' experiences could also bridge the barrier between specialists, the general public, and those working on creative projects to engage and inspire.

ACKNOWLEDGMENTS

This study is part of a doctoral project investigating the role for social robots in healthy independent ageing, and KR is funded by an ESRC Industrial Strategy studentship and ESC is funded by a Starting Grant from the European Research Council [ERC-2015-StG-67727-SOCIAL ROBOTS]. Many thanks to the Social Brain in Action Laboratory - in particular Bishakha Chaudhury for her support with programming the robot.

REFERENCES

[1] Nass, C. and Moon, Y., 2000. Machines and mindlessness: Social responses to computers. *Journal of social issues*, *56*(1), pp.81-103.

[2] Wiese, E., Metta, G., & Wykowska, A. (2017). Robots as intentional agents: using neuroscientific methods to make robots appear more social. *Frontiers in psychology*, *8*, 1663.

[3] Bartneck, C., Rosalia, C., Menges, R. and Deckers, I., 2005. Robot abuse-a limitation of the media equation.

[4] Bartneck, C., Van Der Hoek, M., Mubin, O., & Al Mahmud, A. (2007, March). "Daisy, daisy, give me your answer do!" switching off a robot. In *2007 2nd ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 217-222). IEEE.

[5] Darling, K., Nandy, P. and Breazeal, C., 2015, August. Empathic concern and the effect of stories in human-robot interaction. In *2015 24th IEEE international symposium on robot and human interactive communication (RO-MAN)* (pp. 770-775). IEEE.

[6] Bartneck, C. and Hu, J., 2008. Exploring the abuse of robots. *Interaction Studies*, *9*(3), pp.415-433.