ETHICAL CONSIDERATIONS OF (CONTEXTUALLY) AFFECTIVE ROBOT BEHAVIOUR

A. VAN MARIS*, N. ZOOK, P. CALEB-SOLLY, M. STUDLEY, A. WINFIELD and
S. DOGRAMADZI

*University of the West of England, Frenchay Campus,
Coldharbour Lane, BS16 1QY, Bristol, United Kingdom*
E-mail: anouk.vanmaris@uwe.ac.uk
www.uwe.ac.uk

The use of social robots can bring many benefits, but raises ethical concerns as well. One of these concerns, emotional deception, was investigated in this research. First, affective robot behaviour is validated, followed by a user study to investigate the effect of story context and affective behaviour on user's affect, perception of the robot and acceptance of the robot. Results show that the implemented affective robot behaviours are perceived as intended, and that there is little influence of these affective robot behaviours on people's affect, perception of the robot and acceptance of the robot. It was found that story context has an influence on the was users interpret the emotion of the robot, where a somber context provided in a lower score for happiness. These results raise awareness to practically validate theoretically founded ethical concerns, as these concerns limit the future development and benefits of social robots.

Keywords: Social Robots; Ethics; Affect; Context.

1. Introduction

Research in the use of social robots in daily life settings has greatly increased over the last decade. However, using robots as a replacement of or addition to a human task raises several ethical considerations. These matters have been discussed in the literature on a theoretical level,[1,2] but rarely investigated in practice. It is essential to investigate the theoretically founded ethical claims in practice, since these ethical claims restrict the future development and possible benefits of social robots. One of the ethical concerns raised involves the use of emotional deception in social robots.[3] Deception occurs when false information is communicated that can benefit the communicator,[4] or when no information is communicated at all.[5] It can be either intentional (the deceiver is aware of the fact that a certain

2

feature will raise false expectations) or unintentional (a certain feature of the deceiver causes expectations that the deceiver means to evoke).[5]

When emotive behaviour is used as a means of deception, thus misrepresenting one's emotional state,[6] it is called emotional deception. (Emotional) deception is created when robots are used in assistive settings,[7] since its social behaviour often does not correspond with its actual capabilities. This is a risk, since users may perceive robots differently than intended and raise expectations that cannot be met.

Opinions regarding the question of whether (emotional) deception is acceptable or not are divided. It is perceived as being unethical, as it encourages users in self-deception.[2] However, others are of the opinion that deception is ethically correct if it increases benefits for the deceived,[8] or as long as there is no betrayal of trust.[9] If trust is breached, this may result in a different human-robot interaction (HRI) outcome than intended.[10]

The current study investigates whether emotional deception, a theoretically founded ethical concern, is validated as a concern in practice as well. First, an online survey was used to validate different affective robot behaviours (happy, sad and non-emotive). Next, a user study was run to investigate whether affective robot behaviour (emotional deception) has an influence on the affective state of participants, on their perception of the robot and on their level of acceptance of the robot.

2. Online Survey: Validation of Affective Robot Behaviours

Emotional human-robot interaction can result in the user perceiving the robot as a reliable robot assistant instead of simply a machine for its utility.[11] This online survey investigated whether affective robot behaviour is recognized when no context is provided to support the behaviour. The robot used for this experiment is the social robot Pepper, developed by the company SoftBank Robotics. The survey was distributed through the online Qualtrics survey platform.

2.1. *Robot Behaviour*

Video fragments show the robot saying 'The country Brazil is named after a tree' in three different affective states. This sentence was chosen after piloting, since it had to provide little context to investigate the recognition of the intended affective behaviours. Cues to display different affective behaviours entail body posture and head position,[12] and voice pitch and speed of speech.[11] This experiment distinguishes between sad, happy and

non-emotive behaviour, with different parameters for the characteristics mentioned before.

2.2. *Procedure*

The survey started with information and consent, followed by demographics (age, gender, level of education and familiarity in interacting with robots on a 5-point Likert scale, 1 = not at all familiar, 5 = very familiar). Participants would then see nine short (approx. 4s) video fragments, three for all three affective states. The first fragment depicted non-emotive robot behaviour, the other eight fragments were randomly ordered. Participants had to drag a slider on a scale from sad to happy, depicted in Fig. 1. Each affective state was shown three times to ensure that recognition of the affective state was measured.

Fig. 1.   Five-point scale to rate robot emotion

2.3. *Participants*

Out of 253 participants, 161 people provided usable data (98 male, 61 female, 2 preferred not to say, age range from 18 to 74 years old). Two people did not agree to the consent form, and five did not give permission to use the gathered data after ending the survey. The other 89 participants did not complete the survey.

2.4. *Results*

There was a statistically significant difference in recognized affective behaviour ($F(2,160) = 241.48$, $p < 0.001$). Paired sample t-tests showed the differences between affective behaviours:

- Happy & Non-Emotive ($t(160) = -8.90$, $p < 0.001$)
- Sad & Non-Emotive ($t(160) = 19.86$, $p < 0.001$)
- Sad & Happy ($t(160) = 20.15$, $p < 0.001$)

4

After a Bonferroni correction with a significance level of $p = 0.02$, age significantly influenced perception of the affective robot behaviour ($F(5,155) = 3.30$, $p = 0.007$), where participants between the age of 55/64 on average gave lower ratings than participants of other ages.

- Happy: 18-24 / 55-64 ($p = 0.003$) and 25-34 / 55-64 ($p < 0.001$)
- Non-Emotive: 25-34 / 55-64 ($p = 0.01$)

2.5. *Discussion*

The findings show that people can recognize happy, non-emotive and sad affective robot behaviour, even when no context is provided. Although participants of all ages recognized affective robot behaviour as intended, there were differences between how happy the happy behaviour was rated, similar for non-emotive behaviour. The next study investigates whether adding context while displaying affective robot behaviour influences how the robot is perceived.

3. Face-to-Face Study: Influence of Affective Robot Behaviour

In this study, context was provided during the interaction. It was hypothesized that, due to people's tendency to anthropomorphize, context is sufficient for people to perceive affect in a robot without providing affective robot behaviour. This study used a between-subject design, investigating the effect of story context (cheerful, somber) and affective robot behaviour (happy, sad, non-emotive). The robot tells either a cheerful or a somber story on polar bear cubs. The cheerful story is told either in a happy or non-emotive state. The somber story is told either in a sad or non-emotive state. The materials used in this experiment are similar to the online survey.

3.1. *Measurements*

Questionnaires involved demographics (age, gender, etc.), explicit[13,14] and implicit[15] affect, perception[16] and acceptance.[17] Other questions were the level of familiarity with social robots on a 5-point Likert scale, and interpretation of the robot's emotional state (as used in the online survey).

3.2. *Procedure*

The experiment started with an oral introduction and explanation of the experiment. It was followed by the participants receiving an information sheet

and consent form, and signing them if there were no questions. Before the experiment started, participants were ensured they could terminate the experiment at any time if they felt uncomfortable or did not want to continue. The experiment started with demographics and affect questionnaires, followed by the robot introducing itself and telling a (cheerful or somber) story on polar bear cubs. After that, there were more questionnaires regarding affect, perception of the robot, acceptance of the robot and emotional state of the robot. The session ended with a verbal and written debrief on the goal of the experiment, and how it fits into the larger goal of providing ethical guidelines for the future development of social robots. The robot behaviour for this experiment builds on the results from the online survey used to validate the affective robot behaviours.

3.3. *Participants*

Participants were recruited through the Psychology Participants Pool at the University of the West of England, Bristol in the United Kingdom, receiving one participant credit after completing the experiment. 48 students (age $M = 21.15$, $SD = 2.77$) took part and completed the experiment (4 males per condition, 16 total; 8 females per condition, 32 total). The conditions were:

- cheerful context & happy behaviour
- cheerful context & non-emotive behaviour
- somber context & sad behaviour
- somber context & non-emotive behaviour

3.4. *Results*

3.4.1. *User Affect*

All measures of affect (implicit positive and negative, explicit positive and negative) taken before the interaction correlated strongly with the measures taken after the interaction. There was a strong, positive correlation between the change in explicit positive affect and implicit positive affect ($r = 0.383$, $N = 48$, $p = 0.007$). No other correlations between explicit and implicit change were found.

Explicit negative affect was significantly lower ($t(47) = 7.01$, $p < 0.001$) after interacting with the robot ($M = 15.27$, $SD = 6.82$) than before interacting with the robot ($M = 19.44$, $SD = 7.67$). No other significant differences were found between measures.

6

3.4.2. *Affective Robot Behaviour*

Affective robot behaviour did not influence user affect, whether it was implicit or explicit, or positive or negative. This result was found for story context as well. The results are shown in Table 1 and Table 2.

Table 1. Influence of affective robot behaviour on user affect.

Affect		F	p
Explicit	Positive	0.61	0.55
	Negative	2.15	0.13
Implicit	Positive	0.72	0.49
	Negative	0.015	0.99

Table 2. Influence of story context on user affect.

		F	p
Explicit	Positive	3.89	0.055
	Negative	0.31	0.58
Implicit	Positive	0.94	0.34
	Negative	0.046	0.83

Lastly, Table 3 and Table 4 show that neither affective robot behaviour nor story context had an influence on perception and acceptance of the robot.

Table 3.   Perception of the Robot

	F	p
Affective Behaviour	0.57	0.83
Story Context	0.56	0.73

Table 4.   Acceptance of the Robot

	F	p
Affective Behaviour	0.97	0.52
Story Context	0.95	0.51

3.4.3. *Emotional State of the Robot*

Contradicting the results from the online survey, participants were not always capable of interpreting the emotional states of the robot as intended:

- Happy & Non-Emotive ($t(34) = 0.65$, $p = 0.52$)
- Sad & Non-Emotive ($t(34) = 2.31$, $p = 0.027$)
- Sad & Happy ($t(22) = 2.31$, $p = 0.031$)

This indicates that people were able to distinguish between happy and sad behaviour and non-emotive and sad behaviour. However, there was no significant distinction between happy and non-emotive robot behaviour.

The context of the story did significantly influence how the emotional state of the robot was interpreted ($t(46) = 2.11$, $p = 0.040$), where the robot was rated less happy when the context of the story was somber. However, the robot was rated as slightly happy when the story context was somber ($M = 3.7$, $SD = 0.91$ for somber context, $M = 4.2$, $SD = 0.72$ for cheerful context).

3.5. *Discussion*

Neither the affective behaviour of the robot, nor context of the story significantly impacted participants' affect. However, independent of the condition participants were in, negative affect decreased after interacting with the robot. These findings suggest that interacting with the robot decreased participants' negative affect, and that, due to the absence of other influences, the ethical consequences of affective robot behaviour are limited.

The distinction between affective states in the robot was weaker than in the online survey, as no clear distinction was perceived between happy and non-emotive behaviour. However, the context of the story did significantly influence the perceived emotional state of the robot, with the robot being perceived as more happy when telling the cheerful story.

These findings support the hypothesis that providing context results in people projecting the intended emotions to the robot, without the corresponding affective behaviours being displayed by the robot. This indicates that the use of emotional deception might not be a necessity for a successful human-robot interaction. Further research will show whether this is true for long-term interactions with a robot as well.

4. Conclusions

This paper investigated whether the ethical concerns raised regarding emotional robot deception during human-robot interaction influenced users' mood, perception and acceptance of the robot. Sad, happy and non-emotive affective robot behaviour were validated. Next the impact of story context (with and without affective robot behaviour) on perceived emotional state of the robot was investigated.

The findings from the online survey showed a difference in interpretation of the emotional state of the robot between 18-34 year old people (which included participants of the face-to-face study) and people of age 55+. Therefore, further research will investigate whether results regarding context and affective state are different as well.

8

Affective robot behaviour did not influence the affective state of the participants. This indicates that emotional deception, with regards to the expressiveness of the robot, raises a smaller ethical concern than previously expected and needs to be investigated further. These findings indicate that it is essential to validate theoretically founded ethical concerns through practical research experiments. Further research will investigate the long-term effects of affect and context on older adults, and whether these findings are true for different platforms as well.

Acknowledgements

References

1. K. Dautenhahn and A. Billard, 366 (1999).
2. R. Sparrow, *Ethics and information Technology* **4**, 305 (2002).
3. M. Boden, J. Bryson, D. Caldwell, K. Dautenhahn, L. Edwards, S. Kember, P. Newman, V. Parry, G. Pegman, T. Rodden *et al.*, *Connection Science* **29**, 124 (2017).
4. R. C. Arkin, P. Ulam and A. R. Wagner, *Proceedings of the IEEE* **100**, 571 (2012).
5. A. Dragan, R. Holladay and S. Srinivasa, *Autonomous Robots* **39**, 331 (2015).
6. I. S. Fulmer, B. Barry and D. A. Long, *Journal of Business Ethics* **88**, 691 (2009).
7. A. Sharkey and N. Sharkey, *Ethics and information technology* **14**, 27 (2012).
8. J. Shim and R. C. Arkin, 2328 (2013).
9. A. Matthias, *Kennedy Institute of Ethics Journal* **25**, 169 (2015).
10. P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser and R. Parasuraman, *Human Factors* **53**, 517 (2011).
11. D.-S. Kwon, Y. K. Kwak, J. C. Park, M. J. Chung, E.-S. Jee, K.-S. Park, H.-R. Kim, Y.-M. Kim, J.-C. Park, E. H. Kim *et al.*, 351 (2007).
12. A. Beck, L. Cañamero, A. Hiolle, L. Damiano, P. Cosi, F. Tesser and G. Sommavilla, *International Journal of Social Robotics* **5**, 325 (2013).
13. D. Watson, L. A. Clark and A. Tellegen, *Journal of personality and social psychology* **54**, p. 1063 (1988).
14. D. Watson and L. A. Clark (1999).
15. M. Quirin, M. Kazén and J. Kuhl, *Journal of personality and social psychology* **97**, p. 500 (2009).
16. C. Bartneck, D. Kulić, E. Croft and S. Zoghbi, *International journal of social robotics* **1**, 71 (2009).
17. M. Heerink, B. Kröse, V. Evers and B. Wielinga, *International journal of social robotics* **2**, 361 (2010).