AISB Convention 2015, 20-22nd April, Canterbury

The Society for the Study of Artificial Intelligence and Simulation of Behaviour

AISB Convention 2015
University of Kent, Canterbury

Proceedings of the AISB 2015 Symposium on New Frontiers in Human-Robot Interaction

Edited by Maha Salem, Astrid Weiss, Paul Baxter and Kerstin Dautenhahn

ISBN: 978-1-908187-51-2

Introduction to the Convention

The AISB Convention 2015—the latest in a series of events that have been happening since 1964—was held at the University of Kent, Canterbury, UK in April 2015. Over 120 delegates attended and enjoyed three days of interesting talks and discussions covering a wide range of topics across artificial intelligence and the simulation of behaviour. This proceedings volume contains the papers from the *Symposium on New Frontiers in Human-Robot Interaction,* one of eight symposia held as part of the conference. Many thanks to the convention organisers, the AISB committee, convention delegates, and the many Kent staff and students whose hard work went into making this event a success.

—Colin Johnson, Convention Chair

Introduction to the Symposium

The Symposium on “New Frontiers in Human-Robot Interaction (HRI)” is the fourth of a series of symposia held in conjunction with the AISB convention. Its topics cover cutting-edge interdisciplinary research on understanding, designing, and evaluating robotic systems for and with humans. Its main difference to other HRI-related conferences and workshops is its inclusiveness for exploratory research and the amount of time for open discussion. This year’s symposium consists of six sessions covering topics such as verbal and non-verbal interaction, people’s perception of robots, and ethical issues. Moreover, it includes keynote talks by Mark Coeckelbergh and Angelika Peer and a panel on the topic “Robot Perception and Acceptance”.

Introduction

Human-Robot Interaction (HRI) is a quickly growing and very inter- disciplinary research field. Its application areas will have an impact not only economically, but also on the way we live and the kinds of relationships we may develop with machines. Due to its interdisciplinary nature of the research different views and approaches towards HRI need to be nurtured.

In order to help the field to develop, the Symposium on New Frontiers in Human-Robot Interaction encourages submissions in a variety of categories, thus giving this event a unique character. The symposium consists of paper presentations, panels and, importantly, much time for open discussions which distinguishes this event from regular conferences and workshops in the field of HRI.

History

The first symposium on “New Frontiers in Human-Robot Interaction” was held as part of AISB 2009 in Edinburgh, Scotland; the second symposium was run in conjunction with AISB 2010 in Leicester, England; the third symposium took place during AISB 2014 at Goldsmiths, University of London, England. These three previously organised symposia were characterised by excellent presentations as well as extensive and constructive discussions of the research among the participants. Inspired by the great success of the preceding events and the rapidly evolving field of HRI, the continuation of the symposium series aims to provide a platform to present and discuss collaboratively recent findings and challenges in HRI.

Submission Categories

In order to enable a diverse program, the symposium offers a variety of submission categories, which go beyond typical conference formats. The fourth symposium offered the following categories in the call for papers:

**N* Novel research findings resulting from completed empirical studies.* In this category we encourage submissions where a substantial body of findings has been accumulated based on precise research questions or hypotheses. Such studies are expected to fit within a particular experimental framework (e.g. using qualitative or quantitative evaluation techniques) and the reviewing of such papers apply relevant (statistical and other) criteria accordingly. Findings of such studies should provide novel insights into human-robot interaction studies.

**E* Exploratory studies.* Exploratory studies are often necessary to pilot and fine-tune the methodological approach, procedures and measures. In a young research field such as HRI with novel applications and various robotic platforms, exploratory studies are also often required to derive a set of concrete research questions or hypotheses, in partic- ular concerning issues where there is little related theoretical and experimental work. Although care must be taken in the interpretation of findings from such studies, they highlight issues of great interest and relevance to peers.

**S* Case studies.* Due to the nature of many HRI studies, a large-scale quantitative approach is sometimes neither feasible nor desirable. However, case study evaluation provides meaningful findings if presented appropriately. Thus, case studies with only one participant, or a small group of participants, are encouraged if they are carried out and analysed in sufficient depth.

**P* Position papers.* While categories N, E and S required reporting on HRI studies or experiments, position papers can be conceptual or theoretical, providing new interpretations of known results. Also, in this category we consider papers that present new ideas without having a complete study to report on. Papers in this category are judged on the soundness of the argument presented, the significance of the ideas and the interest to the HRI community.

**R* Replication of HRI studies.* To develop as a field, HRI findings obtained by one research group need to be replicated by other groups. Without any additional novel insights, such work is often not publishable. Within this category, authors have the opportunity to report on studies that confirm or disconfirm findings from experiments that have already been reported in the literature. This category includes studies that report on negative findings.

**D* Live HRI Demonstrations.* Contributors have the opportunity to provide live demonstrations (live or via Skype), pending the outcome of negotiations with the local organisation team. The demo should highlight interesting features and insights into HRI. Purely entertaining demonstrations without significant research content are discouraged.

**Y* System Development.* Research in this category included the design and development of new sensors, robot designs and algorithms for socially interactive robots. Extensive user studies are not necessarily required in this category.

Natural Interaction with Social Robotics

The Fourth Symposium on “New Frontiers in Human-Robot Interaction” was organised in conjunction with the Topic Group on Natural Interaction with Social Robotics. This Topic Group was launched within the EU Horizon 2020 funding framework (http://ec.europa.eu/programmes/horizon2020/), with the strategic goal to keep the topic of interaction prominent in the future calls for European projects. An overview on the list of topics and interests of the Topic Group can be found on the website: http://homepages.stca.herts.ac.uk/~comqkd/TG-NaturalInteractionWithSocialRobots.html.

As the symposium offers an ideal opportunity to discuss related research topics that are relevant for the Topic Group, we introduced one new submission category:

**TG* Topic Group Submissions on “Natural Interaction with Social Robots”.* Submissions in this category will be discussed in a session dedicated to the euRobotics Topic Group “Natural Interaction with Social Robots”. Topics specifically relevant to the TG are e.g. benchmarking of levels of social abilities, multimodal interaction, and human-robot interaction and communication.

Programme Overview

This year’s symposium consists of 17 talks, based on submissions in the following categories:

- *N* Novel research findings resulting from completed empirical studies: 5 submissions
- *E* Exploratory studies: 5 submissions
- *P* Position papers : 4 submissions
- *Y* System Development: 2 submissions
- *TG* Topic Group Submissions on “Natural Interaction with Social Robots”: 1 submission

The talks are structured in six sessions:

1. Ethical issues in HRI
2. Robots' impact on human performance
3. Verbal interaction
4. Facial expressions & emotions
5. Non-verbal cues & behaviours
6. Robot perception & acceptance

The final session is followed by a panel discussion on the same topic. Two invited keynote talks complete the program:

1. Mark Coeckelbergh: "Human-like Robots and Automated Humans: Socializing and Contextualizing HRI"
2. Angelika Peer: "Towards Remote Medical Diagnosticians"

Conclusion

In summary, the symposium mainly focuses on novel empirical findings on human-robot interaction and their impact on our everyday life. Moreover, also theoretical aspects and ethical issues are dis- cussed. We hope these articles show some future research directions for fellow HRI researchers and stimulate ideas for future European projects on natural interaction with social robots.

Contents

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

General Republics' Opinions on Robot Ethics: Comparison between Japan, the USA, Germany, and France

Tatsuya Nomura [1]

Abstract. Ethical issues on robots need to be investigated based on international comparison because general publics' conceptualizations of and feelings toward robots differ due to different situations with respect to mass media and historical influences of technologies. As a preliminary stage of this international comparison, a questionnaire survey based on openended questions was conducted in Japan, the USA, Germany and France ($N = 100$ from each countries). As a result, it was found that (1) people in Japan tended to react to ethical issues of robotics more seriously than those in the other countries, although those in Germany tended not to connect robotics to ethics, (2) people in France tended to specify unemployment as an ethical issue of robotics in comparison with the other countries, (3) people in Japan tended to argue the restriction of using and developing robots as a solution for the ethical problems, although those in France had the opposite trend.

1 Introduction

The recent development of robotics has begun to introduce robots into our daily lives in our homes, schools, and hospitals. In this situation, some philosophers and scientists have been discussing robot ethics [8, 15, 12, 4, 2]. Asaro [1] argued that robot ethics should discuss the following three things: the ethical systems to be built into robots, the ethics of people who design and use robots, and ethical relationships between humans and robots. Lin [6] proposed the following three broad (and interrelated) areas of ethical and social concerns about robotics:

Safety and errors: including mistakes of recognition by battle robots and security against hacking.

Law and ethics: including codes of ethics to be programed into robots, companionships between humans and robots, responsibility of robot behaviors.

Social impact: including economical and psychological change of the society.

Recently, several researchers have been investigating solutions for these ethical problems. However, the opinions of the general public of different countries have not sufficiently been investigated from the perspective of robot ethics. Some existing studies found the general public's preferences of robot types in the context of domestic use [14], expectation of task types in domestic household robots [11], attitudes regarding robots' suitability for a variety of jobs [17], safety perception of humanoid robots [5], and fear and anxiety [9]. However, these survey studies did not focus on the ethical issues of robots.

Moreover, the ethical issues of robots need to be investigated based on international comparison because general publics' conceptualizations of and feelings toward robots differ due to different situations with respect to mass media and historical influences of technologies. In fact, recent studies [16, 19, 13, 18] show differences of opinions of robots between countries, including attitudes toward robots [3, 20], images of robots [10], and implicit attitudes [7]. In addition, interpretations of the word "ethics" differ between countries because of different social norms. Thus, we should compare the opinions of the general publics of several countries when they face the words "robots" and "ethics" at the same time. This comparison will contribute to preparation of discussion on the international consensus of robotics applications.

As a preliminary stage of the international comparison on robot ethics issues, a questionnaire survey based on open–ended questions was conducted in Japan, the USA, and Europe. To take into account the historical influences of wars into the ethical perspectives of military robotics, the survey in Europe was conducted in Germany and France, which were a defeated country and a victorious country in World War II, respectively. This paper reports the results of the survey and then discusses the implications.

2 Method

2.1 Participants and Data Collection Procedure

The survey was conducted from January to February, 2013. Respondents were recruited by a survey company (Rakuten Research). When the survey was conducted, the numbers of possible respondents registered to the company was about 2,300,000 in Japan, 2,780,000 in the USA, 310,000 in Germany, and 450,000 in France. Among the people randomly selected from these large pools of samples based on gender and age, a total of 100 people of ages ranging from 20's to 60's participated in the survey in each of the four countries. Table 1 shows the sample numbers based on country, gender, and age categories.

[1] Department of Media Informatics, Ryukoku University, Japan, email: nomura@rins.ryukoku.ac.jp

The questionnaire consisting of open–ended items was conducted via Internet homepages in all the countries.

Table 1. Sample Numbers Based on Countries, Gender, and Age Categories

		20's	30's	40's	50–60's	Total
Japan	Male	13	12	13	12	50
	Female	12	13	12	13	50
	Total	25	25	25	25	100
USA	Male	11	13	12	14	50
	Female	11	14	18	7	50
	Total	22	27	30	21	100
Germany	Male	12	11	16	11	50
	Female	10	12	15	13	50
	Total	22	23	31	24	100
France	Male	10	15	12	13	50
	Female	20	8	10	12	50
	Total	30	23	22	25	100
Total		99	98	108	95	400

2.2 Measures

As mentioned in the introduction section, the survey aimed at investigating interpretations of the general publics when they face the words robots and ethics at the same time. To measure and compare their primitive conceptualization between the countries, we did not instruct the definitions of "robots" or "ethics".

The questionnaire solicited information about (1) age, (2) gender, (3) occupation (subject of study if respondents were students), and (4) three questions about ethics and robotics. The questionnaire items about ethics and robotics were open–ended, and designed to elicit a wide variety of responses:

Q1: What would you image when hearing "robots" and "ethics" at the same time?

Q2: What sort of ethical problems would happen when robots widespread in society?

Q3: How should we solve the problems mentioned in item 2?

The questionnaire was conducted in Japanese, English, German, and French languages in Japan, the USA, Germany, and France, respectively. The response sentences in Germany and France were translated into English.

3 Results

3.1 Coding of Open–Ended Responses

For quantitative analyses, the open–ended responses were manually classified into categories based on the contents of the responses. This classification coding was determined by two coders. The first coder dealt with both Japanese and English sentences. The second coder consisted of two people, one for the Japanese sentences and another for the English sentences.

First, coding rules were created for each item. Then, two coders independently conducted the coding of 40% of the responses ($N = 40$ from all the responses of each country), and calculated the κ–coefficients showing the degrees of agreement between the two coded results in order to validate the reliability of the coding rules. The coefficients showed sufficient reliability of the coding rules. Table 2 shows coding rule numbers, examples of sentences in the coding, and κ–coefficients. Furthermore, the two coders interactively discussed the contents of the responses and coding results until they reached a consensus about each coding.

3.2 Q1: Images When Hearing "Robots" and "Ethics" at the Same Time

In Q1, each participant's response was classified into one of the three categories shown in Table 2. Responses assigned L0 showed no concrete image. In the German and French samples, several wrote sentences meaning that the words "robots" and "ethics" clashed with each other. Responses assigned L1 stated images from science fiction contents. Responses assigned L2 included realistic concerns of robotics in society and ambiguous apprehension toward the development of robots.

Table 3 shows the distributions of answer categories based on the countries and the results of a χ^2-test and a residual analysis with $\alpha = .05$. Approximately 60% of the respondents mentioned some apprehension toward robotics. The χ^2–test showed differences between the countries in the category distribution. The residual analysis revealed that in the Japan sample, the frequency of L0 was lower than average and that of L1 was higher than average at statistically significant levels. Moreover, in the German samples, the frequency of L0 was higher than average and that of L2 was lower than average. Furthermore, in the French samples, the frequency of L1 was lower than average and that of L2 was higher than average at statistically significant levels.

To visualize the relationships between countries and images of robots and ethics, a correspondence analysis was performed for the cross-table shown in Table 3. The correspondence analysis allows us to visualize the relationship between categories appearing in a cross–table in two-dimensional space. In this visualization, categories similar to each other are placed at proximate positions. Our analysis using this method aims to clarify the relationship between the countries and respondents' images when hearing "robots" and "ethics" at the same time. We should note that the dimensional axes extracted from the data in the cross–table are specific to the table data and are used to visualize relative distances between categories; that is, they do not correspond to any absolute measure, and so it is difficult to assign realistic meanings to these axes.

Figure 1 shows the results of the analysis. The USA is positioned at the middle point between the three answer categories, and Germany is located at L0. Japan is positioned at the middle point between L1 and L2, and France is near L2. These results can be summarized as follows:

- Compared with the other countries, less German respondents specified images in which robots and ethics appeared at the same time.
- More French respondents specified apprehension toward robotics than did the respondents in the other countries.
- More Japanese respondents specified images from virtual contents in comparison with the respondents in the other countries.

Table 2. Coding Rules of Open–Ended Responses and Reliability

Item	Rule	Label		κ
Q1:	R1:	L0:	Responses that did not image any concrete problems (e.g., "nothing", "don't think ...")	.747
		L1:	Responses that mentioned virtual contents including movies, animations, and comics (e.g., "Robocop", "Blade Runner")	
		L2:	Ones except for the above L0 and L1 (e.g., "What are the ethical rules to apply when using robots?"	
Q2:	R21:	L1:	Responses that mentioned unemployment problems (e.g., "Job losses", "Replacing people with robots so unemployment")	.922
		L0:	Others	
	R22:	L1:	Responses that mentioned crimes or wars (e.g., "People use them to spy", "With battle robots, that will make killing easier and easier")	.717
		L0:	Others	
	R23:	L1:	Responses that mentioned some problems except unemployment, crimes and wars (e.g., "Accidents by robots", "There will be no difference between humans/robots")	.711
		L0:	Others	
Q3:	R3:	L0:	Responses that did not mention any concrete problems in Q2	.647
		L1:	Responses that mentioned restriction of robots' functions, methods of using robots, and areas of robot applications, and legal preparation for the restriction (e.g., "Only use robots in certain situations", "Don't give robots the ability of "think"")	
		L2:	Ones except for the above L0 and L1 (e.g., "I have no idea", "Improvement of human morals", "Keep our manual skills")	

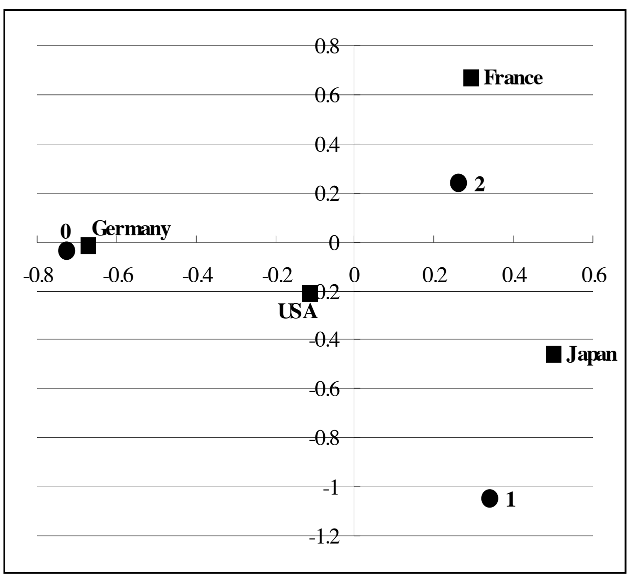


Figure 1. Result of Correspondence Analysis for Table 3

Table 3. Distribution of Answer Categories for Q1 and Results of χ^2–Test and Residual Analysis ($\alpha = .05$)

	Answer Category of R1			
	L0	L1	L2	Total
Japan	$18^{\downarrow}$	$21^{\uparrow}$	61	100
USA	30	15	55	100
Germany	$41^{\uparrow}$	10	$49^{\downarrow}$	100
France	21	$5^{\downarrow}$	$74^{\uparrow}$	100
Total	110 (27.5%)	51 (12.75%)	239 (59.75%)	400 (100%)

$\chi^2(6) = 28.448$, $p < .001$
$^{\uparrow}$: higher than the expected frequency
$^{\downarrow}$: lower than the expected frequency
L0: Responses that did not image any concrete problems
L1: Responses that mentioned virtual contents including movies, animations, and comics
L2: Ones except for the above L0 and L1

3.3 Q2: Ethical Problems in Society

In Q2, one response included several different problems. Thus, each participant's response was assigned multiple labels based on the following rules: (R21) whether it mentioned unemployment problems due to robots, (R22) whether it mentioned the use of robots in crimes and wars, and (R23) whether it mentioned some problems besides unemployment, crimes, and wars. Responses assigned as L1 in R23 included apprehension toward the physical and economical risks of robots, their influences on humans' psychological states, and ambiguous differences between robots and humans.

Table 4 shows the distributions of answer categories based on the countries and the results of the χ^2–test and the residual analysis with $\alpha = .05$. The results can be summarized as follows:

- In the Japan sample, fewer respondents mentioned unem-

Table 4. Distribution of Answer Categories for Q2 and Results of χ^2–Test adn Residual Analysis ($\alpha = .05$)

	R21: Unemployment		R22: Crimes and Wars		R23: Other Problems	
	Not mentioned	Mentioned	Not mentioned	Mentioned	Not mentioned	Mentioned
Japan	$87^{\uparrow}$	$13^{\downarrow}$	$85^{\downarrow}$	$15^{\uparrow}$	$34^{\downarrow}$	$66^{\uparrow}$
USA	77	23	$84^{\downarrow}$	$16^{\uparrow}$	$65^{\uparrow}$	$35^{\downarrow}$
Germany	82	18	$97^{\uparrow}$	$3^{\downarrow}$	47	53
France	$64^{\downarrow}$	$36^{\uparrow}$	$97^{\uparrow}$	$3^{\downarrow}$	$60^{\uparrow}$	$40^{\downarrow}$
Total	310 (77.5%)	90 (22.5%)	363 (90.75%)	37 (9.25%)	206 (51.5%)	194 (48.5%)
	$\chi^2(3) = 16.803$,	$p < .01$	$\chi^2(3) = 18.673$,	$p < .001$	$\chi^2(3) = 23.261$,	$p < .001$

$^{\uparrow}$: higher than the expected frequency, $^{\downarrow}$: lower than the expected frequency

ployment problems at a statistically significant level in comparison with the other countries.

  - More respondents in the French sample mentioned unemployment.

- The respondents mentioning crimes and wars as ethical problems of robotics in society were in the minority (less than 10%).
  - Nevertheless, more respondents mentioned these problems in the Japan and USA samples than in the German and French samples at statistically significant levels.
- More respondents mentioned problems besides unemployment, crimes, and wars in the Japan samples than in the samples of the other countries.
  - On the other hand, fewer respondents in the USA and French samples mentioned these problems than in the Japan and German samples.

3.4 Q3: Solutions for Ethical Problems of Robotics

In Q3, each participant's response was classified into one of the three categories shown in Table 2. Responses assigned label L0 corresponded to the ones that did not specify anything on the ethical problems of robotics in society in Q2 (that is, participants assigned L0 for R21, R22, and R23). In Q3, responses assigned label L1 mentioned restriction of robots functions, methods of using robots, and areas of robot applications. Some responses classified into this category mentioned the need of legal preparation for the restriction. Responses assigned label L2 included the ones that did not provide any concrete solution or the ones that did show some solutions except restriction of robots.

Table 5 shows the distributions of the answer categories based on the countries and the results of the χ^2–test and the residual analysis with $\alpha = .05$. The χ^2–test showed differences between the countries in the category distribution. The residual analysis revealed that in the Japan sample, the frequency of L0 was lower than average and that of L1 was higher than average at statistically significant levels. About half of them mentioned restriction of robotics usage as a solution to their ethical problems. Moreover, it was found that in the German samples, the frequency of L0 was higher than average. Furthermore, in the French samples, the frequency of L1 was lower than average and that of L2 was higher than average at statistically significant levels.

In the same way as Q1, the correspondence analysis for Q3 in Table 5 was conducted to visualize relationships between countries and solution categories for the ethical problems of robots. Figure 2 shows the result. Japan was positioned far from L0 and L2, near L1. France was positioned far from L0 and L1, near L2. The USA and Germany were positioned at the middle of L0 and L1, far from L2. These results can be summarized by the following comparisons between the countries:

- More respondents in Japan specified ethical problems of robots in society and mentioned restriction of robots in terms of functions and methods of usage as a solution to the problems.
- Fewer French respondents mentioned restriction of robots as the problem solution.
- In the USA and particularly in Germany, many respondents did not specify any problem or solution for the ethical issues of robots in society.

Table 5. Distribution of Answer Categories for Q3 and Results of χ^2–Test adn Residual Analysis ($\alpha = .05$)

	Answer Category of R3			
	L0	L1	L2	Total
Japan	$6^{\downarrow}$	$52^{\uparrow}$	42	100
USA	26	43	31	100
Germany	$27^{\uparrow}$	43	30	100
France	21	$30^{\downarrow}$	$49^{\uparrow}$	100
Total	80 (20%)	168 (42%)	152 (38%)	400 (100%)

$\chi^2(6) = 26.536$, $p < .001$
$^{\uparrow}$: higher than the expected frequency
$^{\downarrow}$: lower than the expected frequency
L0: Responses that did not mention any concrete problems in Q2
L1: Responses that mentioned restriction of robots' functions, methods of using robots, and areas of robot applications, and legal preparation for the restriction
L2: Ones except for the above L0 and L1

4 Discussion

4.1 Findings

The survey results suggest some characteristics of Japan, the USA, Germany, and France when the general public of each country faces the issues regarding robot ethics.

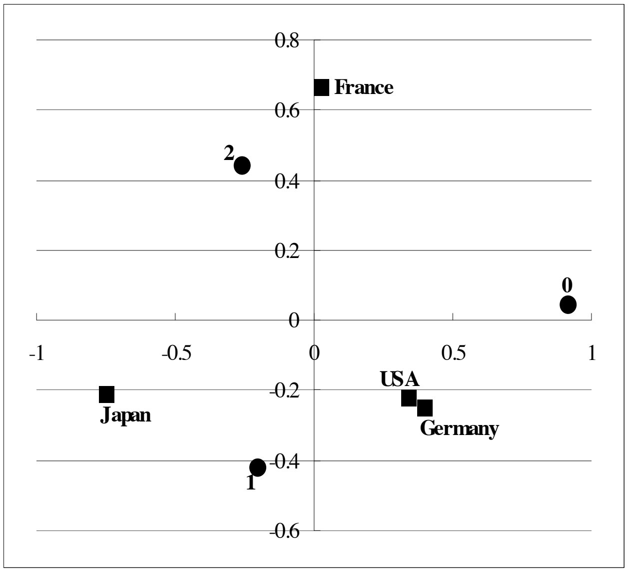


Figure 2. Result of Correspondence Analysis for Table 5

People in Japan tended to react to ethical issues of robotics more seriously than those in the USA, Germany, and France, while they were more influenced by virtual contents such as science fiction movies. In contrast, people in Germany were least likely to connect robotics to ethics. People in France, despite also being in the EU, had a different trend from those in Germany in the sense that they expressed more apprehension toward robotics.

Unemployment as an ethical issue of robotics showed different reactions between these four countries. In particular, Japan and France had opposite trends with respect to this problem. Relationships of robotics with crimes and wars also showed different reactions between the countries. Although a minority of people mentioned this issue as overall, more people tended to specify the issue in Japan and in the USA than in the two European countries.

Consideration of the solutions for the ethical problems of robotics showed opposite trends in Japan and France. Unlike the people in France, the people in Japan tended to argue for restricting the use and development of robots as a solution to ethical problems.

4.2 Implications

The above findings in the survey imply some problems when discussing issues regarding robot ethics at the international level.

First, differences are possible between countries on their general publics awareness of issues regarding robot ethics. Some people may not assume the existence of ethical problems related to robotics. It is implied that the rate of participants in the discussion about robot ethics in society may change depending on the country. Second, it is possible that individual problems have impact on the general public in different ways in different countries. People in one country may participate in discussing an ethical issue and those in another country may not. Such differences in attitudinal biases toward the discussion of robot ethics between countries would make it hard to share problems and solutions internationally. If an ethical problem regarding robots is serious in a country and potentially poses a risk in another country, leaders of the discussion should take into account the differences of awareness of the problem between the countries to establish common assumptions and ways of discussion.

4.3 Limitations

The survey adopted three simple questions and open–ended responses. Thus, the differences of opinions between countries are superficial, and deep factors causing the differences were not explored. It is estimated that these factors include religious beliefs and historical backgrounds in countries, particularly with regard to unemployment and wars. Moreover, the concept of robots may differ between countries [10].

The total number of samples in the survey was not enough to generalize the findings. To clarify more strictly differences in the general publics opinions regarding robot ethics between countries and investigate causes of the differences, we should conduct future surveys using detailed questionnaire items having sufficient validity with a wider area of samples.

ACKNOWLEDGEMENTS

The research was supported in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research No. 21118006 and 25280095.

The author deeply thank Ms. Kanako Tomita and Anderson Li in ATR Intelligent Robotics and Communication Laboratories for their cooperation with the analyses of data in the paper.

REFERENCES

[1] P. M. Asaro, 'What should we want from a robot ethic?', *International Review of Information Ethics*, **6**, 9–16, (2006).

[2] W. Barendregt, A. Paiva, A. Kappas, A. Vasalou, C. Heath, S. Serholt, C. Basedow, and A. O. Patrcia, 'Child–robot interaction: Social bonding, learning and ethics', in *Workshop proceedings of Interaction Design and Children Conference IDC14*, (2014).

[3] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, 'The influence of people's culture and prior experiences with Aibo on their attitude towards robots', *AI & Society*, **21**(1–2), 217–230, (2007).

[4] G. Briggs and M. Scheutz, 'How robots can affect human behavior: Investigating the effects of robotic displays of protest and distress', *International Journal of Social Robotics*, **6**(3), 343–355, (2014).

[5] H. Kamide, Y. Mae, K. Kawabe, S. Shigemi, and T. Arai, 'Effect of human attributes and type of robots on psychological evaluation of humanoids', in *Proc. IEEE Workshop on Advanced Robotics and its Social Impacts*, pp. 40–45, (2012).

[6] P. Lin, 'Introduction to robot ethics', in *Robot Ethics: The ethical and social implications of robotics*, eds., P. Lin, K. Abney, and G. A. Bekey, 3–15, MIT Press, (2011).

[7] K. F. MacDorman, S. K. Vasudevan, and C-C. Ho, 'Does Japan really have robot mania? comparing attitudes by implicit and explicit measures', *AI & Society*, **23**(4), 485–510, (2009).

[8] A. Moon, P. Danielson, and H.F.M. Van der Loos, 'Survey–based discussions on morally contentious applications of interactive robotics', *International Journal of Social Robotics*, **4**(1), 77–96, (2012).

[9] T. Nomura, K. Sugimoto, D. S. Syrdal, and K. Dautenhahn, 'Social acceptance of humanoid robots in japan: A survey for development of the Frankenstein Syndrome Questionnaire', in *Proc. 12th IEEE–RAS International Conference on Humanoid Robots*, pp. 242–247, (2012).

[10] T. Nomura, T. Suzuki, T. Kanda, J. Han, N. Shin, J. Burke, and K. Kato, 'What people assume about humanoid and animal-type robots: Cross-cultural analysis between Japan, Korea, and the USA', *International Journal of Humanoid Robotics*, **5**(1), 25–46, (2008).

[11] L. Oestreicher and K. S. Eklundh, 'User expectations on human–robot co–operation', in *Proc. IEEE International Symposium on Robot and Human Interactive Communication*, pp. 91–96, (2006).

[12] L. D. Riek and D. Howard. A code of ethics for the human–robot interaction profession. Presented at WeRobot 2014 Conference, March 2014.

[13] L. D. Riek, N. Mavridis, S. Antali, N. Darmaki, Z. Ahmed, M. A. Neyadi, and A. Alketheri, 'Ibn sina steps out: Exploring arabic attitudes toward humanoid robots', in *Proc. 36th Annual Convention of the Society for the Study for Artificial Intelligence and the Simulation of Behaviour (AISB 2010)*, volume 1, pp. 88–94, (2010).

[14] M. Scopelliti, M. V. Giuliani, and F. Fornara, 'Robots in a domestic setting: A psychological approach', *Universal Access in the Information Society*, **4**(2), 146–155, (2005).

[15] A. Sharkey and N. Sharkey, 'Granny and the robots: ethical issues in robot care for the elderly', *Ethics and Information Technology*, **14**(1), 27–40, (2012).

[16] T. Shibata, K. Wada, Y. Ikeda, and S. Sabanovic, 'Cross–cultural studies on subjective evaluation of a seal robot', *Advanced Robotics*, **23**(4), 443–458, (2009).

[17] L. Takayama, W. Ju, and C Nass, 'Beyond dirty, dangerous and dull: What everyday people think robots should do', in *Proc. 3rd ACM/IEEE International Conference on Human-Robot Interaction*, pp. 25–32, (2008).

[18] G. Trovato, M. Zecca, S. Sessa, L. Jamone, J. Ham, K. Hashimoto, and A. Takanishi, 'Cross–cultural study on human-robot greeting interaction: acceptance and discomfort by egyptians and japanese', *Paladyn, Journal of Behavioral Robotics*, **4**(2), 83–93, (2013).

[19] S. Šabanović, 'Robots in society, society in robots', *International Journal of Social Robotics*, **2**(4), 439–45–, (2010).

[20] L. Wang, P-L. P. Rau, V. Evers, B. Krisper, and P. Hinds, 'When in Rome: The role of culture & context in adherence to robot recommendations', in *Proc. 5th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 359–366, (2010).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Differences on Social Acceptance of Humanoid Robots between Japan and the UK

Tatsuya Nomura[1], Dag Sverre Syrdal[2], and Kerstin Dautenhahn[2]

Abstract. To validate a questionnaire for measuring people's acceptance of humanoid robots in cross-cultural research (the Frankenstein Syndrome Questionnaire: FSQ), an online survey was conducted in both the UK and Japan including items on perceptions of the relation to the family and commitment to religions, and negative attitudes toward robots (the NARS). The results suggested that 1) the correlations between the FSQ subscale scores and NARS were sufficient, 2) the UK people felt more negative toward humanoid robots than did the Japanese people, 3) young UK people had more expectation for humanoid robots, 4) relationships between social acceptance of humanoid robots and negative attitudes toward robots in general were different between the nations and generations, and 5) there were no correlations between the FSQ subscale scores, and perception of the relation to the family and commitment to religions.

1 INTRODUCTION

In recent years, several studies have revealed the influences of human cultures into feelings and behaviors toward robots [1, 2, 3, 4, 5, 6], and some of them focused on social acceptance of robots. Evers, et al. [1] revealed differences between the US and Chinese people on their attitudes toward and the extent to which they accepted choices made by a robot. Li, et al. [2] found an interaction effect between human cultures (Chinese, Korean and German) and robots' tasks (teaching, guide, entertainment and security guard) on their engagement with the robots. Yueh and Lin [5] showed differences on preferences of home service robots between Taiwanese and Japanese people.

The research group also have been developing a questionnaire to measure and compare humans' acceptance of humanoid robots between nations, and explore factors influencing social acceptance of humanoids including cultural ones [7, 8]. The questionnaire, called "Frankenstein Syndrome Questionnaire" (FSQ), aims at clarification of differences on social acceptance of humanoid robots between the Westerners and Japanese based on Kaplan's idea [9] reflecting the concept of "Frankenstein Syndrome" originated from genetic engineering [10]. The surveys using this questionnaire suggested age differences on acceptance of humanoid robots in Japan [11], and some differences between the UK and Japan [8].

However, the previous studies had some problems on sampling in the sense that data from an online survey and that based on a normal paper-and-pencil method were mixed in one nation sample. As a result, the factor structure extracted from the sample was not stable [12]. Moreover, the previous survey did not take into account verification of criterion-related validity of the questionnaire.

To overcome the above problems, an online survey was conducted in both the UK and Japan under more strict control of sampling. The survey included another psychological scale of which validity had already been supported, the Negative Attitudes toward Robots Scale [13]. The scale was used to verify correlations between social acceptance of humanoid robots and attitudes toward robots in general, to investigate the criterion-related validity of the Frankenstein Syndrome Questionnaire.

As well as cultures, the survey aimed at exploring other factors related to social acceptance of humanoid robots. As factors to be explored, the survey firstly focused on age. In the survey conducted in Japan about ten years ago, our research group found that persons in their 40s had positive opinions of robots in comparison with other generations [14]. Thus, the survey aimed at comparing one group of persons in their 50s with another in their 20s to clarify age differences. Moreover, a survey conducted in Japan and Sweden adopted perceptions of the relation to the family and commitment to religions as indices reflecting differences between these different nations [15]. Thus, the survey also included these two factors "the relation to the family" and "commitment to religions".

The paper reports the results of the survey, and discusses the implications from the perspective of development of humanoid robots.

2 Method

2.1 Date and Participants:

The survey was conducted from January to February 2014. 100 Japanese and 100 UK respondents were recruited by a survey company at which about one million and six hundred thousand Japanese and one million and one hundred thousand UK persons have registered. Respondents in each nation were limited to people who were born and had been living only in the corresponding nation. The respondents consisted of fifty persons in their 20s (male: 25, female: 25) and fifty persons in their 50s (male: 25, female: 25) in each of the nations.

The homepage of the online survey had been open for these participants during the above period. The questionnaire of the online survey was conducted with the native language for the respondents in each of the nations.

2.2 Survey Design:

The questionnaire did not give the explicit definition of robots, or include any photo and image of robots, except for the instruction on humanoid robots just before conducting the Frankenstein Syndrome Questionnaire. The scale on attitudes

[1] Dept. of Media Informatics, Ryukoku Univ., Shiga 520-2194, Japan. Email: nomura@rins.ryukoku.ac.jp.
[2] Adaptive Systems Research Group, Univ. of Hertfordshire, AL10 9AB, UK. Email: {D.S.Syrdal, K.Dautenhahn}@herts.ac.uk.

toward robots in general was firstly conducted, and then the Frankenstein Syndrome Questionnaire was conducted since the reverse order had a possibility that envisions of humanoids evoked by the conduction of the FSQ affected the measurement of attitudes toward robots in general. The concrete items and scales in the survey were as follows:

Perception of the Relation to the Family and Commitment to Religions:

The following two items, which were used in the comparison survey between Japan and the Northern Europe by Otsuka et al. [15], were presented on the face sheet measure participants' degrees of perception of the relation to the family and commitment to religions:

- Do you think you relate to your family members? (five-graded answer from "1. I completely agree" to "5. I completely disagree")
- Does such notion as "I have nothing to do with religion or faith" apply to you? (five-graded answer from "1. It strongly applies to me" to "5. It does not apply to me at all.")

Negative Attitudes toward Robots Scale (NARS):

To measure participants' attitudes toward robots in general, the NARS [13] was adopted in the survey. The scale consists of 14 items classified into three subscales. The first subscale (S1, six items) measures negative attitude toward interaction with robots (e.g., "I would feel paranoid talking with a robot."). The second subscale (S2, 5 items) measures negative attitude toward the social influence of robots (e.g., "Something bad might happen if robots developed into living beings."). The third subscale (S3, 3 items) measures negative attitude toward emotional interaction with robots (e.g., "I feel comforted being with robots that have emotions.").

Each item is scored on a five-point scale: 1) strongly disagree; 2) disagree; 3) undecided; 4) agree; 5) strongly agree, and an individual's score on each subscale is calculated by adding the scores of all items included in the subscale, with some items reverse coded.

Frankenstein Syndrome Questionnaire (FSQ):

The questionnaire was developed to measure acceptance of humanoid robots including expectations and anxieties toward this technology in the general public [8,11]. It consists of 30 items shown in Table 1. Each questionnaire item was assigned with a seven-choice answer (1: "Strongly disagree", 2: "Disagree", 3: "Disagree a little", 4: "Not decidable", 5: "Agree a little", 6: "Agree", 7: "Strongly agree".).

Just before conducting the FSQ, the definition of "humanoids robots" was instructed only with texts as follows:

"Humanoid robots are robots that roughly look like humans, that have two arms, legs, a head, etc. These robots may be very human-like in appearance (including details such as hair, artificial skin etc.), but can also have machine-like features (such as wheels, a metal skin etc)."

3 RESULTS

3.1 Subscales of the FSQ and Reliability:

Although previous studies had explored the factor structures in the FSQ [8,13], they were sufficiently not stable to be replicated across studies [12]. To extract the subscales of the FSQ again, a factor analysis with maximum likelihood method and Promax rotation was conducted for the 30 items. Although the analysis found five factors having eigen values more than 1, the scree plot showed that the difference on the eigen values between the fourth and fifth factors was small. Thus, the factor analysis was conducted based on four-factor structure. The cumulative contribution of these four factors was 52.8%.

After removing items having factor loadings more than .3 on more than one item, item analysis using Cronbach's α-coefficients and I-T correlations was performed for each factor in turn to select items in the corresponding subscale. Table 1 shows the results of these analyses.

The subscale corresponding to the first factor consisted of 9 items representing negative feelings toward social impacts of humanoid robots such as "Humanoid robots may make us even lazier." Thus, the subscale was interpreted as "negative feelings toward humanoid robots." The subscale corresponding to the second factor consisted of 8 items representing positive expectation of humanoid robots in the society such as "Humanoid robots can be very useful for teaching young kids." Thus, the subscale was interpreted as "expectation for humanoid robots". The subscale corresponding to the third factor consisted of 3 items representing negative feelings toward humanoid robots at religious and philosophical levels such as "The development of humanoid robots is blasphemous." Thus, the subscale was interpreted as "root anxiety toward humanoid robots". The fourth factor was removed in the analysis since it consisted of only two items.

Cronbach's reliability coefficients α, showing the internal consistencies of the subscales, were .899 for "negative feelings toward humanoid robots," .861 for "expectation for humanoid robots," and .859 for "root anxiety toward humanoid robots." These values showed sufficient internal consistencies for all three subscales. The score of each subscale was calculated as the sum of the scores of all items included in the subscale ("negative feelings toward humanoid robots": max 63, min 9, "expectation for humanoid robots": max 56, min 8, and "root anxiety toward humanoid robots": max 21, min 3).

3.2 Comparison between Nations and Generations:

FSQ Subscale Scores:

Three-way ANOVAs with gender by nation (Japan vs. UK) by generation (20's vs. 50's) were conducted for the subscale scores of the FSQ. Table 2 shows the results. For "negative feelings toward humanoid robots," the main effects of gender and nations were at statistically significant levels although the effect size on gender was small. For "expectation for humanoid robots," only the first order interaction effect between nations and generations was at a statistically significant level.

Figure 1 shows the means and standard deviations of the subscale scores of "negative feelings toward humanoid robots" and "expectation for humanoid robots". Bonfferoni Post Hoc tests revealed that the UK respondents in their 20s had higher expectation for humanoid robots than the UK respondents in

Item No.	Item Sentences	Factor I	II	III	IV
30	Widespread use of humanoid robots would take away jobs from people.	**.929**	.076	-.098	-.212
4	Humanoid robots may make us even lazier.	**.766**	.037	-.057	-.077
12	If humanoid robots cause accidents or trouble, persons and organizations related to development of them should give sufficient compensation to the victims.	**.705**	.113	-.285	.132
8	I am afraid that humanoid robots will encourage less interaction between humans.	**.697**	.026	.167	-.015
20	I feel that if we become over-dependent on humanoid robots, something bad might happen.	**.681**	-.071	-.011	.245
17	I would hate the idea of robots or artificial intelligences making judgments about things.	**.655**	-.132	.279	-.045
11	I would feel uneasy if humanoid robots really had emotions or independent thoughts.	**.548**	-.055	-.004	.178
27	Something bad might happen if humanoid robots developed into human beings.	**.512**	-.048	.191	.193
23	*Humanoid robots should perform dangerous tasks, for example in disaster areas, deep sea, and space.*	***.493***	***.346***	*-.242*	*.055*
16	I am concerned that humanoid robots would be a bad influence on children.	**.491**	-.171	.245	.148
24	*Many humanoid robots in society will make it less warm.*	***.452***	*.009*	***.396***	*.144*
13	I can trust persons and organizations related to development of humanoid robots.	-.147	**.777**	.256	-.018
15	Humanoid robots can be very useful for teaching young kids.	-.225	**.737**	.262	.077
10	I don't know why, but I like the idea of humanoid robots.	-.259	**.733**	.044	.295
25	*I trust persons and organizations related to the development of humanoid robots to disclose sufficient information to the public, including negative information.*	*-.015*	***.720***	***.314***	*-.210*
19	Humanoid robots can make our lives easier.	.204	**.672**	-.282	.118
3	Persons and organizations related to development of humanoid robots are well-meaning.	.103	**.672**	-.018	-.054
18	Humanoid robots are a natural product of our civilization.	-.072	**.660**	.083	-.111
28	Persons and organizations related to development of humanoid robots will consider the needs, thoughts and feelings of their users.	.303	**.547**	-.119	.022
5	Humanoid robots can be very useful for caring the elderly and disabled.	.054	**.544**	-.184	.144
6	Humanoid robots should perform repetitive and boring routine tasks instead of leaving them to people.	.123	**.524**	-.053	.200
29	The development of humanoid robots is blasphemous.	-.032	.013	**.892**	.001
9	The development of humanoid robots is a blasphemy against nature.	-.038	.000	**.863**	.077
26	*Technologies needed for the development of humanoid robots belong to scientific fields that humans should not study.*	*-.072*	*.203*	***.663***	*.058*
21	I don't know why, but humanoid robots scare me.	.297	-.205	**.567**	.006
22	*I feel that in the future, society will be dominated by humanoid robots.*	***.314***	***.331***	***.403***	*-.186*
1	*I am afraid that humanoid robots will make us forget what it is like to be human.*	*.234*	*-.097*	***.379***	***.323***
7	People interacting with humanoid robots could sometimes lead to problems in relationships between people.	.240	.049	.292	**.547**
2	*Humanoid robots can create new forms of interactions both between humans and between humans and machines.*	*.010*	***.433***	*-.112*	***.474***
14	Widespread use of humanoid robots would mean that it would be costly for us to maintain them.	.248	.099	.037	**.452**

(Items shown with Italic: reduced based on the criterion of factor loadings more than .3 on more than one item and item analysis)

Table 1. Items of the Frankenstein Syndrome Questionnaire and Results of Factor Analysis

their 50s ($p < .0.13$) and the Japan participants in their 20's ($p < .0.55$). There were neither main effects nor any interactions for "root anxiety toward humanoid robots" (mean = 9.9, *SD* = 4.1).

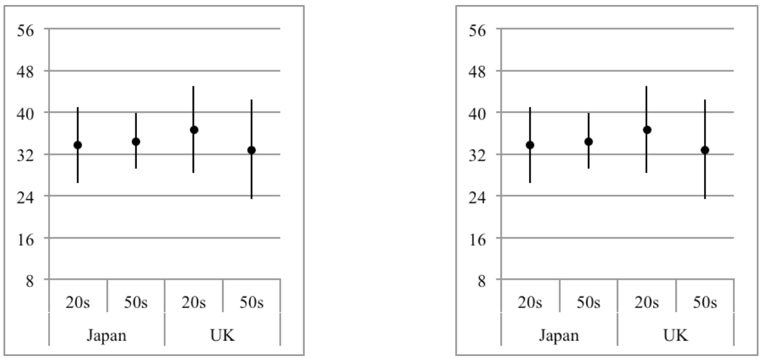
<u>Correlations with the NARS, Perception of the Relation to the Family, and Commitment to Religions:</u>

The Cronbach's α-coefficients for the NARS subscales were .854, .779, and .842 for S1, S2, and S3, respectively. These values showed that these subscales had sufficient internal consistency.

Table 3 shows Pearson's correlation coefficients between the FSQ subscale scores, the NARS subscale scores, and item scores of relation to family and religious commitment based on the nations and generations. Tests of equality on correlation coefficients found statistically significant differences between the four respondents groups, suggesting the following trends:

		Main Effect			First Order Interaction			Second Order Interaction
		Gender	Nation	Generation	Gender X Nation	Gender X Generation	Nation and Generation	
I. Negative Feelings toward Humanoid Robots	*F*	6.121	24.630	.406	.027	.444	2.420	.985
	p	.014	< .001	.525	.871	.506	.121	.322
	η^2	.027	.108	.002	.000	.002	.011	.004
II. Expectation for Humanoid Robots	*F*	2.281	.376	2.013	.185	3.186	4.548	.855
	p	.133	.540	.158	.668	.076	.034	.356
	η^2	.011	.002	.010	.001	.016	.022	.004
III. Root Anxiety toward Humanoid Robots	*F*	1.877	.676	2.702	1.606	1.437	.264	.019
	p	.172	.412	.102	.207	.232	.608	.891
	η^2	.009	.003	.013	.008	.007	.001	.000

Table 2. Results of ANOVAs for the FSQ Subscale Scores



Negative Feelings Expectation

Figure 1. Means and Standard Deviations of Scores of Negative Feelings toward and Expectation for Humanoid Robots

- Between "negative feelings toward humanoid robots" and "expectation for humanoid robots" ($\chi^2(3)$ = 19.677, *p* < .001): positive correlation in the Japan respondents in their 20s, and negative correlation in the UK respondents in their 50s,
- Between "negative feelings toward humanoid robots" and "negative attitude toward social influences of robots" ($\chi^2(3)$ = 11.091, *p* < .05): moderate levels of correlations in the respondents in their 20s, and strong correlations in the respondents in their 50s,
- Between "negative feelings toward humanoid robots" and "negative attitude toward emotional interaction with robots" ($\chi^2(3)$ = 14.468, *p* < .01): moderate levels of positive correlations only in the respondents in their 50s,
- Between "expectation for humanoid robots" and "root anxiety toward humanoid robots" ($\chi^2(3)$ = 12.840, *p* < .01): a moderate level of negative correlation only in the UK respondents in their 50s,
- Between "expectation for humanoid robots" and "negative attitude toward social influences of robots" ($\chi^2(3)$ = 13.715, *p* < .01): moderate levels of negative correlations only in the respondents in their 50s,
- Between "root anxiety toward humanoid robots" and "expectation for humanoid robots" ($\chi^2(3)$ = 11.770, *p* < .01): strong correlation in the Japan respondents in their 20's, and moderate levels of correlations in the other respondents,
- Between "root anxiety toward humanoid robots" and "negative attitude toward emotional interaction with robots" ($\chi^2(3)$ = 8.279, *p* < .05): a moderate level of positive correlation only in the UK respondents in their 50s.

On the other hand, there were moderate levels of positive correlations between "negative feelings toward humanoid robots" and "root anxiety toward humanoid robots", between "negative feelings toward humanoid robots" and "negative attitude toward interaction with robots", and between "root anxiety toward humanoid robots" and "negative attitude toward interaction with robots". Moreover, there was a moderate level of negative correlation between "expectation for humanoid robots" and "negative attitude toward social influences of robots".

There were no correlations between the FSQ subscale scores, and perception of the relation to the family and commitment to religions, although only the UK participants in 50's showed statistically significant correlations between these scores and perception of the relation to the family.

4. DISCUSSION

4.1 Findings:

The survey results suggest sufficient correlations between the FSQ subscale scores and NARS. It supports the criterion-related validity of the FSQ. Negative attitude toward interaction with

		FSQII	FSQIII	NARSS1	NARSS2	NARSS3	Religion	Family
FSQI	Whole	-.059	.472**	.426**	.664**	.139	.012	-.081
	Jp 20s	.381**	.534**	.316*	.605**	-.117	.001	-.179
	Jp 50s	-.234	.617**	.431**	.744**	.411**	.143	.196
	UK 20s	.149	.474**	.446**	.478**	-.049	-.133	-.147
	UK 50s	-.402**	.431**	.516**	.820**	.461**	.121	.223
FSQII	Whole		-.208**	-.076	-.169*	-.554**	-.095	-.182**
	Jp 20s		.125	.008	.186	-.383**	.047	-.155
	Jp 50s		-.182	-.159	-.307*	-.473**	-.022	-.157
	UK 20s		-.195	-.037	-.064	-.698**	-.247	-.007
	UK 50s		-.544**	-.261	-.487**	-.584**	-.079	-.317*
FSQIII	Whole			.620**	.526**	.089	.034	.054
	Jp 20s			.734**	.757**	-.113	-.113	-.101
	Jp 50s			.604**	.391**	.191	.034	.233
	UK 20s			.588**	.345*	.020	.124	-.070
	UK 50s			.562**	.593**	.420**	.138	.308*

FSQI: Negative Feelings toward Humanoid Robots, FSQII: Expectation for Humanoid Robots,
FSQIII: Root Anxiety toward Humanoid Robots,
NARSS1: Negative Attitude toward Interaction with Robots, NARSS2: Negative Attitude toward Social Influences of Robots,
NARSS3: Negative Attitude toward Emotional Interaction with Robots,
Religion: Religious Commitment, Family:Relation to Family

Table 3. Pearson's Correlation Coefficients between FSQ and NARS Subscale Scores, and Item Scores of Relation to Family and Religious Commitment

robots in general was related to negative feelings and root anxiety toward humanoid robots in both the UK and Japan.

The survey results also suggest some differences on social acceptance of humanoid robots between the two countries. The UK participants felt more negative towards humanoid robots than their Japanese counterparts. In addition, the UK participants in their 20s had more positive expectations for humanoid robots than any other group..

These results suggest some differences dependent on generation, on relationships between social acceptance of humanoid robots and negative attitudes toward robots in general. The correlation between negative attitudes toward emotional interaction with robots and negative feelings toward humanoids was at a moderate level only in 50s people. The correlation between negative attitude toward social influences of robots and expectation for humanoids also had the similar trend. The correlation between negative attitude toward emotional interaction with robots and root anxiety toward humanoids was at a moderate level only in UK participants in their 50s.

4.2 Implications:

The results in the survey imply that people in the UK have more negative feelings toward humanoid robots than those in Japan. This however, depends on the generation of the participants. Likewise, relationships between feelings toward humanoid robots and attitudes toward robots in general also depend on the generation of respondent. This suggests that changing attitudes toward some particular types of robots may not lead to acceptance of other types of robots, nor robots in general.

In order to further social acceptance of humanoid robots across cultures, designers of robots need to consider individual, generational, and cultural factors in their potential users.

4.3 Limitations and Future Works:

The survey did not take into account concrete attitudes toward the relation to family and religious commitment. It may lead to non-correlation between these factors and social acceptance of robots. On the other hand, previous research has found correlations between these factors and negative attitudes toward robots [16]. It suggests that religious and family factors may indirectly influence social acceptance of humanoid robots. Future surveys need to include this indirect influence in the survey design.

Moreover, the survey did not adopt any image stimulus of robots in order to avoid influences of images of specific types of robots. Future surveys should include more sophisticated items while exploring dominant images of robots in the corresponding nations.

In addition, the survey did not consider possible differences between human attitudes toward humanoid robots measured in questionnaires and live interactions with them, such as dealt with by Wang, et al. [17]. We need to conduct experiments to investigate how psychological constructs measured by the FSQ affect human behaviors toward humanoid robots in real situations.

REFERENCES

[1] V. Evers, H. Maldonado, T. Brodecki, P. Hinds, 'Relational vs. group self-construal: Untangling the role of national culture in HRI', in Proc. ACM/IEEE International Conference on Human-Robot Interaction, pp.255-262, (2009).

[2] D. Li, P. L. P. Rau, and Y. Li. 'A Cross-cultural Study: Effect of Robot Appearance and Task', International Journal of Social Robotics, 2(2), 175-186, (2010).

[3] G. Eresha, M. Haring, B. Endrass, E. Andre, and M. Obaid, 'Investigating the influence of culture on proxemic behaviors for humanoid robots', in Proc. IEEE International Symposium on Robot and Human Interactive Communication, pp.430-435, (2013).

[4] M. Makatchev, R. Simmons, M. Sakr, and M. Ziadee, ‘Expressing ethnicity through behaviors of a robot character’, in Proc. ACM/IEEE International Conference on Human-Robot Interaction, pp.357-364, (2013).

[5] H-P. Yueh, and W. Lin, ‘The Interaction between Human and the Home Service Robot on a Daily Life Cycle’, in Proc. International Conference on Cross-Cultural Design, pp.175-181, (2013).

[6] Y. Ho, E. Sato-Shimokawara, T. Yamaguchi, and N. Tagawa, ‘Interaction robot system considering culture differences’, in Proc. IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp.170-174, (2013).

[7] D. S. Syrdal, T. Nomura, H. Hirai, and K. Dautenhahn, ‘Examining the Frankenstein Syndrome: An Open-Ended Cross-Cultural Survey’, in Social Robotics: Third International Conference, ICSR 2011, Proceedings (B. Mutlu, et al. (Eds)), pp.125-134, Springer, (2011).

[8] D. S. Syrdal, T. Nomura, and K. Dautenhahn, ‘The Frankenstein Syndrome Questionnaire - Results from a Quantitative Cross-Cultural Survey’, in Social Robotics: 5th International Conference, ICSR 2013 (G. Herrmann, et al. (Eds)), pp.270-279, Springer, (2013).

[9] F. Kaplan, ‘Who is afraid of the humanoid?: Investigating cultural differences in the acceptance of robots’, International Journal of Humanoid Robot., 1(3), 465–480, (2004).

[10] B. E. Rollin, The Frankenstein Syndrome: Ethical and Social Issues in the Genetic Engineering of Animals, Cambridge University Press, (1995).

[11] T. Nomura, K. Sugimoto, D. S. Syrdal, and K. Dautenhahn, ‘Social Acceptance of Humanoid Robots in Japan: A Survey for Development of the Frankenstein Syndrome Questionnaire’, in Proc. IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012), pp.242-247, (2012).

[12] T. Nomura, D. S. Syrdal, and K. Dautenhahn, ‘Cultural Difference on Factor Structure in a Questionnaire about Humanoid Robots’, presented in Workshop on Culture Aware Robotics, May, (2014).

[13] T. Nomura, T. Suzuki, T. Kanda, and K. Kato, ‘Measurement of Negative Attitudes toward Robots’, Interaction Studies, 7(3), 437-454, (2006).

[14] T. Nomura, T. Tasaki, T. Kanda, M. Shiomi, H. Ishiguro, and N. Hagita, ‘Questionnaire-Based Social Research on Opinions of Japanese Visitors for Communication Robots at an Exhibition’, AI and Society, 21(1-2), 167-183, (2007).

[15] M. Otsuka, M. Akiyama, K. Mori, and H. Hoshino, ‘Comparative Study of Values, Work Ethics, and Lifestyles in Japan and Sweden: An Initial Report’, Bulletin of Human Science, 33, 105-119, (2011) (in Japanese).

[16] T. Nomura, ‘Comparison on Negative Attitude toward Robots and Related Factors between Japan and the UK’, ACM International Conference on Collaboration Across Boundaries: Culture, Distance & Technology, pp.87-90, (2014).

[17] L. Wang, P-L. P. Rau, V. Evers, B. Krisper, and P. Hinds, ‘When in Rome: The role of culture & context in adherence to robot recommendations’, in Proc. ACM/IEEE International Conference on Human-Robot Interaction, pp. 359–366, (2010).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Presence of Life-Like Robot Expressions Influences Children's Enjoyment of Human-Robot Interactions in the Field

David Cameron[1], Samuel Fernando[2], Emily Collins[1], Abigail Millings[1], Roger Moore[2], Amanda Sharkey[2], Vanessa Evers[3], and Tony Prescott[1]

Abstract. Emotions, and emotional expression, have a broad influence on the interactions we have with others and are thus a key factor to consider in developing social robots. As part of a collaborative EU project, this study examined the impact of life-like affective facial expressions, in the humanoid robot Zeno, on children's behavior and attitudes towards the robot. Results indicate that robot expressions have mixed effects depending on the gender of the participant. Male participants showed a positive affective response, and indicated greater liking towards the robot, when it made positive and negative affective facial expressions during an interactive game, when compared to the same robot with a neutral expression. Female participants showed no marked difference across two conditions. This is the first study to demonstrate an effect of life-like emotional expression on children's behavior in the field. We discuss the broader implications of these findings in terms of gender differences in HRI, noting the importance of the gender appearance of the robot (in this case, male) and in relation to the overall strategy of the project to advance the understanding of how interactions with expressive robots could lead to task-appropriate symbiotic relationships.

1 INTRODUCTION

A key challenge in human robot interaction (HRI) is the development of robots that can successfully engage with people. Effective social engagement requires robots to present engaging personalities [1] and to dynamically respond to and shape their interactions to meet human user needs [2].

The current project seeks to develop a biologically grounded [3] robotic system capable of meeting these requirements in the form of a socially-engaging *Synthetic Tutoring Assistant* (STA). In developing the STA, we aim to further the understanding of human-robot symbiotic interaction where symbiosis is defined as the capacity of the robot, and the person, to mutually influence each other in a positive way. Symbiosis, in a social context, requires that the robot can interpret, and be responsive to, the behavior and state of the person, and adapt its own actions appropriately. By applying methods from social psychology we aim to uncover key factors in robot personality, behavior, and appearance that can promote symbiosis. We hope that this work will also contribute to a broader theory of human-robot bonding that we are developing drawing on comparisons with our psychological understanding of human-human, human-animal and human-object bonds [4].

A key factor in social interaction is the experience of emotions [5]. Emotions provide important information and context to social events and dynamically influence how interactions unfold over time [6]. Emotions can promote cooperative and collaborative behavior and can exist as shared experiences, bringing individuals closer together [7]. Communication of emotion can be thought of as a request for others to acknowledge and respond to our concerns and to shape their behaviors to align with our motives [8]. Thus emotional expression can be important to dyadic interactions, such as that between a teacher and student, where there is a need to align goals.

Research with a range of robot platforms has demonstrated the willingness of humans to interpret robot expressive behavior – gesture [9], posture [10], and facial expression [1] – as affective communication. The extent to which robot expression will promote symbiosis will depend, however, on how well the use of expression is tuned to the ongoing interaction. Inappropriate use of affective expression could disrupt communication and be detrimental to symbiosis. Good timing and sending clear signals is obviously important.

Facial expression is a fundamental component of human emotional communication [11]. Emotion expressed through the face is also considered to be especially important as a means for communicating evaluations and appraisals [12]. Given the importance of facial expressions to the communication of human affect, they should also have significant potential as a communication means for robots [13]. This intuition has lead to the development of many robot platforms with the capacity to produce human-like facial expression, ranging from the more iconic/cartoon-like [e.g., 14, 15] to the more natural/realistic [e.g., 16, 17, 18].

Given the need to communicate clearly it has been argued that, for facial expression, iconic/cartoon-like expressive robots may be more appropriate for some HRI applications, for instance, where the goal is to communicate/engage with children [16, 15]. Nevertheless, as the technology for constructing robot faces has become more sophisticated, robots are emerging with richly-expressive life-like faces [16, 17, 18], with potential for use in a range of real-world applications including use with children. The current study arose out of a desire to evaluate one side of this symbiotic interaction – exploring the value of life-like facial expression in synthetic tutoring assistants for children. Whilst it is clear that people can distinguish robot expressions almost as well

[1] Dept. of Psychology, University of Sheffield, S10 2TN, UK.
Email: {d.s.cameron, e.c,.collins, a.millings, t.j.prescott}@sheffield.ac.uk
[2] Dept. of Computer Science, University of Sheffield S10 2TN, UK.
Email: {s.fernando, r.k.moore, a.sharkey}@sheffield.ac.uk
[3] Dept. of Electrical Engineering, Mathematics and Computer Science, University of Twente, NL. Email: v.evers@utwente.nl

as human ones [16, 18], there is little direct evidence to show a positive benefit of life-like expression on social interaction or bonding. Although children playing with an expressive robot are more expressive than those playing alone [19], this finding could be a result of the robot's social presence [20] and not simply due to its use of expression. A useful step toward improving our understanding would be the controlled use of emotional expression in a setting in which other factors, such as the presence of the robot and its physical and behavioral design, are strictly controlled.

In the current study the primary manipulation was to turn on or off the presence of appropriate positive and negative facial expressions during a game-playing interaction, with other features such as the nature and duration of the game, and the robot's bodily and verbal expression held constant. As our platform we employed a Hanson Robokind Zeno R50 [21] which has a realistic silicon rubber ("flubber") face, that can be reconfigured, by multiple concealed motors, to display a range of reasonably life-like facial expressions in real-time (Figure 1).

Figure 1. The Hanson Robokind Zeno R50 Robot with example facial expressions

By recording participants (with parental consent), and through questionnaires, we obtained measures of proximity, human emotional facial expression, and reported affect. We hypothesized that children would respond to the presence of facial expression by (a) reducing their distance from the robot, b) showing greater positive facial expression themselves during the interaction, and c) reporting greater enjoyment of the interaction compared to peers who interacted with the same robot but in the absence of facial expression. Previous studies have shown some influence of demographics such as age and gender on HRI [22, 23, 24]. In our study, a gender difference could also arise due to the visual appearance of the Zeno robot as similar to a male child, which could prompt different responses in male and female children. We therefore considered these other factors as potential moderators of children's responses to the presence or absence of robot emotional expression.

2 METHOD

2.1 Design

Due to the potential of repeated robot exposure prejudicing participants' affective responses, we employed a between-subjects design, such that participants were allocated to either the experimental condition – interaction with a facially expressive robot, or to the control condition of a non-facially-expressive robot. Allocation to condition was not random, but determined by logistics due to the real-world setting of the research. The study took place as part of a two-day special exhibit demonstrating modern robotics at a museum in the UK. Robot expressiveness was manipulated between the two consecutive days, such that visitors who participated in the study on the first day were allocated to the expressive condition, and visitors who participated in the study on the second day were allocated to the non-expressive condition.

2.2 Participants

Children visiting the exhibit were invited to participate in the study by playing a game with Zeno. Sixty children took part in the study in total (37 male and 23 female; M age = 7.57, SD = 2.80). Data were trimmed by age to ensure sufficient cognitive capacity (those aged < 5 were excluded[4]) and interest in the game (those aged >11 were excluded) leaving 46 children (28 male and 18 Female; M age = 8.04, SD = 1.93).

2.3 Measures

Our primary dependent variables were interpersonal responses to Zeno measured through two objective measures: affective expressions and interpersonal distance. Additional measures comprised of a self-report questionnaire, completed by participating children, with help from their parent/carer if required, and an observer's questionnaire, completed by parents/carers.

2.3.1 Objective Measures

Interpersonal distance between the child and the robot over the duration of the game was recorded, using a Microsoft Kinect sensor, and mean interpersonal distance during the game calculated. Participant expressions were recorded throughout the game and automatically coded for discrete facial expressions: Neutral, Happy, Sad, Angry, Surprised, Scared, and Disgusted, using Noldus FaceReader version 5. Mean intensity of the seven facial expressions across the duration of the game were calculated. Participants' game performances (final scores) were also recorded. FaceReader offers automated coding of expressions at an accuracy comparable to trained raters of expression [25].

2.3.2 Questionnaires

Participants completed a brief questionnaire on their enjoyment of the game and their beliefs about the extent to which they thought that the robot liked them. Enjoyment of playing Simon Says with Zeno was recorded using a single-item, four-point measure, ranging from 'I definitely did not enjoy it' to 'I really enjoyed it'. Participants' perceptions of the extent to which Zeno liked them single-item on a thermometer scale, ranging from 'I do not think he liked me very much' to 'I think he liked me a lot'. They were also asked if they would like to play the game again. Parents and

[4] Additional reasons for excluding children below the age of 5 were questionable levels of understanding when completing the self-report questionnaires, and low reliability in FaceReader's detection of expressions in young children.

carers completed a brief questionnaire on their perceptions of their child's enjoyment and engagement with the game on single-item thermometer scales, ranging from 'Did not enjoy the game at all' to 'Enjoyed the game very much and 'Not at all engaged' to 'Completely engaged'.

2.4 Procedure

The experiment took place in a publicly accessible lab and prospective participants could view games already underway. Brief information concerning the experiment was provided to parents or carers and informed consent was obtained from parents or carers prior to participation.

During the game, children were free to position themselves relative to Zeno within a 'play zone' boundary marked on the floor by a mat (to delineate the area in which the system would correctly detect movements) and could leave the game at their choosing. The designated play zone was marked by three foam .62msq mats. The closest edge of the play zone was 1.80m from the robot and the play zone extended to 3.66m away. These limits approximate the 'social distance' classification [26]. This range was chosen for 2 reasons i) Participants would likely expect the game used to occur within social rather than public- or personal-distance ii) This enabled reliable recordings of movement by the Kinect sensor. The mean overall distance for the participants from the robot fell well within social-distance boundaries (2.48m).

At the end of the game, participants completed the self-report questionnaire, while parents completed the observer's questionnaire. Participant-experimenter interaction consistency was maintained over the two days by using the same experimenter on all occasions for all tasks.

Interaction with the robot took the form of the widely known *Simon Says* game (Figure 2). This game was chosen for several reasons: children's familiarity with the game, its uncluttered structure allows autonomous instruction and feedback delivery by Zeno, and its record of successful use in a prior field study [27].

The experiment began with autonomous instructions delivered by Zeno as soon as children stepped into the designated play zone in front of the Kinect sensor. Zeno introduced the game by saying, "*Hello. Are you ready to play with me? Let's play Simon Says. If I say Simon Says you must do the action. Otherwise you must keep still*." The robot would then play ten rounds of the game or play until the child chose to leave the designated play zone. In each round, Zeno gave one of three simple action instructions: 'Wave your hands', 'Put your hands up' or 'Jump up and down'. Each instruction was given either with the prefix of 'Simon says' or no prefix.

Figure 2. A child playing Simon Says with Zeno

The OpenNI/Kinect skeleton tracking system was used to determine if the child had performed the correct action in three seconds following instruction. For the 'Wave your hands' action, our system monitored the speed of the hands moving. If sufficient movement for the arms were detected following instruction then the movement was marked as a wave. For the 'Jump up and down' action the vertical velocity of the head was monitored, again with a threshold to determine if a jump had taken place. Finally for the 'Put your hands up' action, our system monitored the positions of the hands relative to the waist. If the hands were found to be above the waist for more than half of the three seconds following the instruction then the action was judged to have been executed. The thresholds for the action detection were determined by previous trial and error during pilot testing in a university laboratory. The resulting methods of action detection were found to be over 98% accurate in our study. In the rare cases where the child did the correct action and the system judged incorrectly then the experimenters would step in and say "*Sorry, the robot made a mistake there, you got it right*".

If children followed the action instruction after hearing 'Simon says' the robot would say, "*Well done, you got that right*". If the child remained still when the prefix was not given, Zeno would congratulate them on their correct action with "*Well done, I did not say Simon Says and you kept still*". Conversely, if the child did not complete the requested movement when the prefix was given Zeno would say, "*Oh dear, I said Simon Says, you should have waved your hands*". If they completed the requested movement in the absence of the prefix, Zeno would inform them of their mistake with, "*Oh dear, I did not say Simon Says, you should have kept still*". Zeno gave children feedback of a running total of their score at the end of each round (the number of correct turns completed).

If the child left the play zone before ten rounds were played, the robot would say, "*Are you going? You can play up to ten rounds. Stay on the mat to keep playing*". The system would then wait three seconds before announcing, "*Goodbye. Your final score was (score)*". This short buffer was to prevent the game ending abruptly if the child accidentally left the play zone for a few seconds.

At the end of the ten rounds, the robot would say, "*All right, we had ten goes. I had fun playing with you, but it is time for me to play with someone else now. Goodbye*."

The sole experimental manipulation coincided with Zeno's spoken feedback to the children after each turn. In the expressive robot condition, Zeno responded with appropriate 'happiness' or 'sadness' expressions, following children's correct or incorrect responses. These expressions were prebuilt animations, provided with the Zeno robot, named 'victory' and 'disappointment' respectively. These animations were edited to remove gestures so only facial expression were present. In contrast, in the non-expressive robot condition, Zeno's expressions remained in a neutral state regardless of child performance. Previous work indicates that children can recognize these facial expression representations by the Zeno robot with a good degree of accuracy [28].

3 RESULTS

A preliminary check was run to ensure even distribution of participants to expressive and non-expressive conditions. There were 9 female and 16 male participants in the expressive condition and 9 female and 12 male participants in the non-expressive condition. A chi square test was run before analysis to check for even gender distribution across conditions indicates no significant difference (X2 (1,48) = 2.25, p = .635).

3.1 Objective Measures

Overall, we did not observe any significant main effects of Zeno's expressiveness on objective measures of interpersonal distance or facial expressions between conditions. However, there were significant interaction effects, when gender was included as a variable.

There was a significant interaction of experimental condition and child's gender on average child's expressions of happiness $F(1,39) = 4.75$, $p = .038$. While male participants showed greater average happiness in the expressive robot condition in comparison to those in the non-expressive condition (19.1%, SE 3.3% versus 5.3%, SE 4.1%), female participants did not differ between conditions (7.4%, SE 4.3% versus 12.6%, SE 4.6%). Simple effects tests (with Bonferroni correction) indicated that the observed differences between conditions for male participants was significant (p = .012).

A contrasting interaction was found for average expressions of surprise $F(1,39) = 5.16$, $p = .029$. Male participants in the expressive robot condition showed less surprise than those in the non-expressive condition (6.1%, SE 3.2% versus 19.6%, SE 4.0%), whereas female participant expressions for surprise did not differ between conditions (11.9%, SE 4.2% versus 7.1%, SE 4.5%). There were no further significant interactions for any of the remaining expressions.

There was a near significant interaction for experimental condition and child's gender for interpersonal distance $F(1,41) = 2.81$, $p = .10$ (Figure 3). Male participants interacting with the expressive robot tended to stand closer (M = 2.28m, SE .10m) than did those interacting with the non-expressive robot (M = 2.57m, SE .13m), whereas female participants interacting with the expressive robot tended to stand further away (M = 2.59m, SE .14m) than those interacting with the non-expressive robot (M = 2.45m, SE .14m). A follow-up simple effect test indicates that the difference between conditions for male participants was also near significant (p = .086).

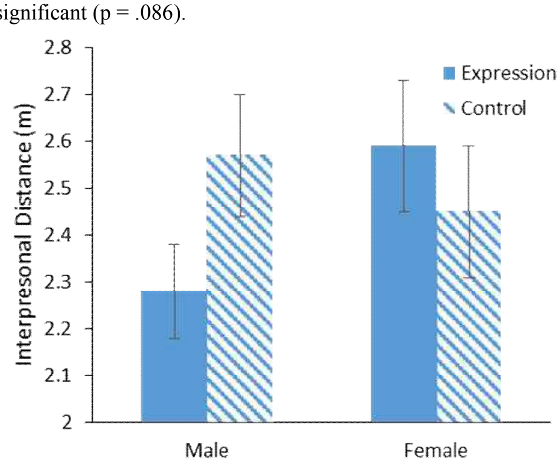


Figure 3. Mean interpersonal distance during game

Controlling for participant age or success/failure in the game made no material difference to any of the objective measures findings.

3.2 Questionnaires

No significant main effects of condition were seen for self-reported measures or observer reported measures. However, there were significant gender effects, and significant gender X condition effects. Gender had a main effect on children's beliefs about the extent to which the robot liked them $F(1,38) = 5.53$, $p = 0.03$. Female participants reported significantly lower ratings (M = 3.08, SE .34) than did male participants (M = 4.17, SE .31).

We observed a significant interaction of gender and experimental condition for participants' enjoyment in interacting with Zeno $F(1,38) = 4.64$, $p = .04$. Male participants interacting with the expressive Zeno reported greater enjoyment of the interaction than those who interacted with the non-expressive Zeno (M = 3.40, SE .18 versus M = 3.00, SE .23), whereas female participants interacting with the expressive Zeno reported less enjoyment than those interacting with the non-expressive Zeno (M = 3.22, SE .23 versus M = 3.78, SE .23). Simple effects tests did not indicate that the difference found between conditions were significant for either male participants (p > .10) or female participants (p > .10).

Results from the observer reports generated by the participants' parents or carers showed the same trends as those from the self-report results but did not show significant main or interaction effects. Controlling for participant age or success/failure in the game made no material difference to any of the questionnaire data findings.

4 DISCUSSION

The results provide new evidence that life-like facial expressions in humanoid robots can impact on children's experience and enjoyment of HRI. Moreover, our results are consistent across multiple modalities of measurement. The presence of expressions could be seen to cause differences in approach behaviors, positive expression, and self-reports of enjoyment. However, the findings are not universal as boys showed more favorable behaviors and views towards the expressive robot compared to the non-expressive robot, whereas girls tended to show the opposite.

Sex differences towards facially expressive robots during HRI could have profound impact on the design and development of future robots; it is important to replicate these experimental conditions and explore these results in more depth in order to identify why these results arise. At this stage, the mechanisms underpinning these differences still remain to be determined. We outline two potential processes that could explain our results.

The current results could be due to children's same-sex preferences for friends and playmates typically exhibited at the ages range tested (ages five to ten) [29]. Zeno is nominally a 'boy' robot and expressions may be emphasizing cues seen on the face to encourage user perceptions of it as a boy. As a result, children may be acting in accordance with existing preferences for play partners [30]. If this is the case, it would be anticipated that replication of the current study with a 'girl' robot counterpart would produce results contrasting with the current findings.

Alternatively, results could be due to the robot's expressions emphasizing the existing social situation experienced by the children. The current study took place in a publically accessible space, with participants in the company of museum visitors, other volunteers, and the children's parents or carer. Results from the current study could represent children's behavior towards the robot based on existing gender driven behavioral attitudes. Girls may have felt more uncomfortable than boys when in front of their parents whilst engaging in explorative play [20] with a strange person (in the form of their perceived proximity to the experimenter) and an unfamiliar object (the robot). Social cues from an expressive robot, absent in a neutral robot, may reinforce these differences through heightening the social nature of the experiment.

Behavioral gender differences in children engaging in public or explorative play are well established, and the link between these gender differences and the influence of direct parents/carers differential socialization of their children dependent upon the sex [31,32], is a further established link of developmental study. To better explore the gender difference observed in our study we must take into consideration existing observed behavioral patterns in children engaging in explorative play around their parents. Replication in a familiar environment away from an audience including children's parents may then impact on apparent sex differences observed in the current HRI study.

The current study is a small-sample field experiment. As with the nature of field studies, maintaining an exacting control over experimental conditions is prohibitively difficult. Along with possible confounds from the public testing space, the primary experimenter knew the condition each child was assigned to; despite best efforts in maintaining impartiality, the current study design cannot rule out potential unconscious experimenter influence on children's behaviors. In studies concerning emotion and expression, potential contagion effects of expression and emotion [33] could impact on participant's expressions and reported emotions. The current results therefore offer a strong indication of the areas to be further explored under stricter experimental conditions.

We aim to repeat the current study in a more controlled experimental environment. Children will complete the same Simon-says game in the familiar environment of their school, this time without an audience. Rather than allocation by day to condition, the study protocol will be modified to randomly allocate children to conditions, and the study will be conducted by an experimenter naïve to conditions. Testing at local schools offers better controls over participant sample demographics as children can be recruited based on age and having similar educational and social backgrounds. The environment of this study also removes any direct influence by the presence of parents/carers. Thus, a repeat of the current study under stricter conditions also offers opportunity to further test the proposed hypotheses for the observed sex differences in enjoyment in interacting with a facially expressive robot.

We have previously proposed that human-robot bonds could be analyzed in terms of their similarities to different types of existing bond with other human, animals, and objects [4]. Our relationships with robots that are lacking in human-like faces may have interesting similarities to human-animal bonds which can be simpler than those with other people—expectations are clearer, demands are lower, and loyalty is less prone to change. Robots with more human-like faces and behavior, on the other hand, may prompt responses from users that include more of the social complexities of human-human interaction. Thus, aspects of appearance that indicate gender can become more important, subtleties of facial and vocal expression may be subjected to greater scrutiny and interpretation. Overall, as we progress towards more realistic human-like robots we should bear in mind that whilst the potential is there for a richer expressive vocabulary, the bar may also be higher for getting the communication right.

5 CONCLUSION

This paper offers further steps towards developing a theoretical understanding of symbiotic interactions between humans and robots. The production of emulated emotional communication through facial expression by robots is identified as a central factor in shaping human attitudes and behaviors during HRI. Results from both self-repot and objective measures of behavior point towards possible sex differences in responses to facially expressive robots; follow-up work to examine these is identified. These findings highlight important considerations to be made in the future development of a socially engaging robot.

6 ACKNOWLEDGMENTS

This work is supported by the European Union Seventh Framework Programme (FP7-ICT-2013-10) under grant agreement no. 611971. We wish to acknowledge the contribution of all project partners to the ideas investigated in this study.

7 REFERENCES

[1] Breazeal, C., & Scassellati, B. How to build robots that make friends and influence people. In *Intelligent Robots and Systems, 1999. IROS'99. Proceedings*. 1999 IEEE/RSJ International Conference on (Vol. 2, pp. 858-863). IEEE.

[2] Pitsch, K., Kuzuoka, H., Suzuki, Y., Sussenbach, L., Luff, P., & Heath, C. "The first five seconds": Contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI. *In Robot and Human Interactive Communication*, 2009. RO-MAN 2009. The 18th IEEE International Symposium on (Toyama, Japan, Sept 27, 2009) IEEE 985-991 DOI:10.1109/ROMAN.2009.5326167

[3] Verschure, P. F. Distributed adaptive control: a theory of the mind, brain, body nexus. *Biologically Inspired Cognitive Architectures*, **1**, 55-72, (2012). DOI:10.1016/j.bica.2012.04.005

[4] Collins, E. C., Millings, A., Prescott, T. J. 2013. Attachment in assistive technology: A new conceptualisation. In *Assistive Technology: From Research to Practice*, Encarnação, P., Azevedo,L., Gelderblom, G. J., Newell, A., & Mathiassen N. IOS Press, 823-828. DOI:10.3233/978-1-61499-304-9-823

[5] Van Kleef, G. A. How emotions regulate social life the emotions as social information (EASI) model. *Current Directions in Psychological Science*, **18**, 184-188, (2009). DOI:10.1111/j.1467-8721.2009.01633.x

[6] Hareli, S., & Rafaeli, A. Emotion cycles: On the social influence of emotion in organizations. *Research in organizational behavior*, **28**, 35-59, (2008). DOI:10.1016/j.riob.2008.04.007

[7] Kelly, J. R., & Barsade, S. G. Mood and emotions in small groups and work teams. *Organizational behavior and human decision processes*, **86**, 99-130, (2001). DOI:10.1006/obhd.2001.2974

[8] Parkinson, B. Do facial movements express emotions or communicate motives. *Personality and Social Psychology Review*, **9**, 278-311, (2005). DOI = 10.1207/s15327957pspr0904_1

[9] Tielman, M., Neerincx, M., Meyer, J., & Looije, R. Adaptive emotional expression in robot-child interaction. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, (Bielefeld, Germany, Mar. 03 – 06, 2014) ACM, New York, NY, 407-414. DOI:10.1145/2559636.2559663.

[10] Beck, A., Cañamero, L., Damiano, L., Sommavilla, G., Tesser, F., & Cosi, P. Children interpretation of emotional body language displayed by a robot. *Social Robotics*, 62–70. (2011) Springer, Berlin Heidelberg.

[11] Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. Communication of affect through facial expressions in humans.

Journal of Personality and Social Psychology, **23**, 362-371, (1972). DOI:10.1037/h0033171
[12] Parkinson, B. Emotions are social. *British Journal of Psychology*, **87**, 663-683, (1996). DOI:10.1111/j.2044-8295.1996.tb02615.x
[13] Nitsch, V., & Popp, M. Emotions in robot psychology. *Biological cybernetics*, 1-9, (2014). DOI: 10.1007/s00422-014-0594-6
[14] Breazeal, C. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, **59**, 119-155, (2003). DOI:10.1016/S1071-5819(03)00018-1
[15] Espinoza, R.R , Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., ... & Pozzi, C. Child-robot interaction in the wild: advice to the aspiring experimenter. In *Proceedings of the 13th international conference on multimodal interfaces* (Alicante, Spain, Nov. 14 – 18, 2011) ACM, New York, NY, 335-342. DOI:10.1145/2070481.2070545
[16] Becker-Asano, C., & Ishiguro, H. Evaluating facial displays of emotion for the android robot Geminoid F. In *Affective Computational Intelligence* (WACI), 2011 IEEE Workshop on (Paris, France, April, 11-15, 2011) IEEE 1-8 DOI:10.1109/WACI.2011.5953147
[17] Fagot, B. I. The influence of sex of child on parental reactions to toddler children. *Child development*, **2**, 459-465, (1978). DOI:jstor.org/stable/1128711
[18] Mazzei, D., Lazzeri, N., Hanson, D., & De Rossi, D. HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. In *Biomedical Robotics and Biomechatronics* (BioRob), 2012 4th IEEE RAS & EMBS International Conference on (Rome, Italy, Jun. 24 -27, 2012) IEEE 195-200 DOI:10.1109/BioRob.2012.6290687
[19] Shahid, S., Krahmer, E., & Swerts, M. Child–robot interaction across cultures: How does playing a game with a social robot compare to playing a game alone or with a friend? *Computers in Human behaviour*, **40**, 86-100, (2014). DOI:10.1016/j.chb.2014.07.043.
[20] Kraut, R. E., & Johnston, R. E. Social and emotional messages of smiling: An ethological approach*. Journal of Personality and Social Psychology*, **37**, 1539-1553, (1979). DOI:10.1037/0022-3514.37.9.1539
[21] Hanson, D., Baurmann, S., Riccio, T., Margolin, R., Dockins, T., Tavares, M., & Carpenter, K.. Zeno: A cognitive character. *AI Magazine*, 9-11, (2009).
[22] Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. Interactive robots as social partners and peer tutors for children: A field trial. *Human-computer interaction*, **19**, 61-84, (2004). DOI:10.1207/s15327051hci1901&2_4
[23] Kuo, I. H., Rabindran, J. M., Broadbent, E., Lee, Y. I., Kerse, N., Stafford, R. M. Q., & MacDonald, B. A. Age and gender factors in user acceptance of healthcare robots. In *Robot and Human Interactive Communication*, 2009. RO-MAN 2009. The 18th IEEE International Symposium on (Toyama, Japan, Sept. 27- Oct. 2, 2009) IEEE. 214-219 DOI:10.1109/ROMAN.2009.5326292
[24] Shahid, S., Krahmer, E., Swerts, M., & Mubin, O. Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia. November 22-26, 2010) ACM, New York, USA. 332-335 DOI:10.1145/1952222.1952294
[25] Lewinski, P., den Uyl, T. M., & Butler, C. Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience, Psychology, and Economics*, **7** 227-236, (2014). DOI:10.1037/npe0000028
[26] Burgess, J. Interpersonal spacing behavior between surrounding nearest neighbors reflects both familiarity and environmental density. *Ethology and sociobiology*, **4**, 11-17, (1983). doi:10.1016/0162-3095(83)90003-1
[27] Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., & Blow, M. KASPAR–a minimally expressive humanoid robot for human–robot interaction research. *Applied Bionics and Biomechanics*, **6**, 369-397, (2009). DOI:10.1080/11762320903123567
[28] Costa, S., Soares, F., & Santos, C. Facial Expressions and Gestures to Convey Emotions with a Humanoid Robot. *In Social Robotics* 542-551). Springer International Publishing. (2013). DOI:10.1007/978-3-319-02675-6_54
[29] Martin, C. L., & Fabes, R. A. The stability and consequences of young children's same-sex peer interactions. *Developmental psychology*, **37**, 431-446, (2001). DOI:10.1037/0012-1649.37.3.431.
[30] Lindsey, E. W. Physical activity play and preschool children's peer acceptance: Distinctions between rough-and-tumble and exercise play. *Early Education and Development*, **25**, 277-294, (2014). DOI:10.1080/10409289.2014.890854
[31] Gonzalez, A. M. *Parenting Preschoolers with Disruptive Behavior Disorders: Does Child Gender Matter*? Dissertation. Washington (2013) University in St. Louis, St Louis, Missouri, USA.
[32] Kim, H. J., Arnold, D. H., Fisher, P. H., & Zeljo, A. Parenting and pre-schoolers' symptoms as a function of child gender and SES. *Child & Family Behavior Therapy*, **27**, 23-41, (2005). DOI: 10.1300/J019v27n02_03
[33] Hatfield, E., Cacioppo, J. T. & Rapson, R. L. *Emotional Contagion.* Cambridge university press, Cambridge, UK, 1994.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

The Paro robot seal as a social mediator for healthy users

Natalie Wood[1] and **Amanda Sharkey**[2] and **Gail Mountain**[3] and **Abigail Millings**[4]

Abstract. Robots are being designed to provide companionship, but there is some concern that they could lead to a reduction in human contact for vulnerable populations. However, some field data suggests that robots may have a social mediation effect in human-human interactions. This study examined social mediation effects in a controlled laboratory setting. In this study 114 unacquainted female volunteers were put in pairs and randomised to interact together with an active Paro, an inactive Paro, or a dinosaur toy robot. Each pair was invited to evaluate and interact with the robot together during a ten minute session. Post-interaction questionnaires measured the quality of dyadic interaction between participants during the session. Our results indicate that the strongest social mediation effect was from the active Paro.

1 INTRODUCTION

Over the last decade robots have been developed as an alternative to companion animals for older-aged adults and people with dementia in care homes. These companion robots are designed to improve the physical and psychological health of users by calming them, providing companionship, and have the potential to help reduce loneliness and improve the well-being of their users [11, 2].

Despite the benefits these assistive robots bring, there are objections to their use with vulnerable populations. Sparrow and Sparrow [15] raise one main concern as the loss of human contact had by these populations as their human carers are replaced with robotic counterparts. They argue that robotic technology is not currently capable of meeting the social and emotional needs of their users. As the amount of human-human contact between patients and their carers decreases, this could lead to a reduction in the number and quality of their social relationships, and therefore their quality of life.

This concern is supported by Sharkey and Sharkey [13], who consider the negative effects of reduced social contact on the physical and psychological well-being of the elderly. They propose that access to human social contact must be considered before robotic technology is brought into elder-care.

However, a recent developing area of research has shown that robotics can have a role in improving human-human relationships. This small but growing body of field data suggests that a companion robot, the Paro robot seal, can be used to encourage social interaction between individuals, in addition to providing human-robot companionship.

The majority of these studies examined the social mediation effect of Paro using samples of people with cognitive impairment in care home settings.

This paper aims to contribute to this research by investigating whether the social mediation effect is present in healthy populations and under controlled conditions. Animals have been found to act as a social catalyst for healthy individuals as well as for people with dementia and older adults [5][9]. We propose that the same could be true of animal-like robots. Our study looks at the ability of Paro to mediate social interaction between strangers by providing an ice breaker effect in a controlled laboratory setting.

Section 1.1 of this paper introduces the existing work on social mediation with Paro. Section 2 details our hypotheses. This is followed by the methodology used for the study in section 3. Our analytic strategy and results are discussed in section 4. We discuss our findings and limitations of the work in section 5. Finally section 6 concludes the paper.

1.1 Background

Previous studies conducted in care homes have reported the ability of Paro as a social mediator. A randomised controlled trial by Robinson, Macdonald, Kerse, and Broadbent [12] showed a significant decrease in the loneliness reported by 17 residents of a retirement home after 12 weeks of regular activity with Paro. They also found an increase in social interaction between residents when they engaged in activity with Paro compared to during normal activities with and without the resident dog.

Wada and Shibata [19] found that the social network of 12 elderly residents in a care home increased after Paro was available in an open public space for two months.

In an ethnographic case study, Giusti and Marti [4] found that not only did the amount of social interaction increase, but the social dynamic between three residents of a nursing home changed from primarily one-to-one social interactions to group interaction involving all three during interactions with Paro.

Kidd, Taggart and Turkle [7] investigated the effect that a small number of interactions with Paro had on social activity in the nursing home setting. They found that the 23 residents reported more social interaction with others when they were with active Paro than when it was turned off. They also found that presence of more people, including caregivers and experimenters, improved the amount of social engagement.

These findings were supported in another nursing home where Šabanović et al. [18] observed that the social interactions increased between seven residents, including those who were not directly interacting with Paro, during robot-assisted therapy sessions.

[1] Sheffield Robotics, University of Sheffield, S1 3JD, UK. Email: natalie.wood@shef.ac.uk.
[2] Department of Computer Science, University of Sheffield, S1 4DP, UK. Email: a.sharkey@dcs.shef.ac.uk.
[3] School of Health and Related Research, University of Sheffield, S1 4DA, UK. Email: g.a.mountain@sheffield.ac.uk
[4] Centre for Assistive Technology & Connected Healthcare, and Department of Psychology, University of Sheffield, S10 2TP, UK. Email: a.millings@sheffield.ac.uk.

Although the results of these studies show support for Paro as a social mediator in the nursing home setting, they are limited by small sample sizes. In addition, the majority of these studies lack control conditions, such that the social meditation effect cannot be attributed specifically to the Paro. It is unclear whether any novel, robotic stimuli would produce the effects observed. In the current study, we examine the social mediation effect of an active Paro which is turned on and interactive, compared to that of an inactive Paro which is turned off and resembling a cuddly toy, and another interactive robotic toy, Pleo the dinosaur.

2 HYPOTHESES

This study aims to answer the following questions: Can the social mediation effect of Paro apply to a healthy population? Can the effect be measured under a controlled laboratory setting?

To investigate the social mediation effect of Paro we invited pairs of strangers to interact for the first time together, along with an active Paro, an inactive Paro, or a Pleo.

We anticipate that the social mediation effect of Paro when active will lead to participants enjoying interacting with the other participant more and having a better experience when interacting together, than with an inactive Paro and the Pleo. We also anticipate that interacting together with an active Paro will lead to a more positive opinion of the other participant compared to the other two conditions.

Secondary to this we also expect the Pleo to be a more effective social mediator than an inactive Paro. This leads to our hypotheses: Primary hypotheses:

- H1: Compared to the Pleo and inactive Paro conditions, the participants in the active Paro condition will report a:
  - (a): higher quality of interaction.
  - (b): higher opinion of the other participant.

Secondary hypotheses:

- H2: Compared to the inactive Paro condition, the participants in the Pleo condition will report a:
  - (a): higher quality of interaction.
  - (b): higher opinion of the other participant.

3 METHODOLOGY

3.1 Participants

Participants were recruited using a number of methods. Firstly, undergraduate psychology students were invited to participate through the University's research participation scheme in exchange for course credit. Secondly, an email was sent using volunteer mailing lists for University of Sheffield staff and students, inviting volunteers to participate in exchange for entry into a prize draw for one of two £30 Amazon vouchers. Female participants were chosen due to the availability of volunteers at the university which were predominantly female at the time.

In total 114 participants were recruited, aged from 15 to 59 ($M = 23.94, SD = 8.38$), and were paired according to availability. Pairs of participants were randomly allocated into conditions with 21 participant pairs in the active Paro condition, 19 participant pairs in the inactive Paro condition, and 17 participants pairs in the Pleo condition.

3.2 Materials

3.2.1 Paro

The Paro was developed in Japan by Shibata [21] as a therapeutic tool for use with people with dementia. It is a pet-like robot based on a harp seal pup and its body is covered in soft, white, and antibacterial fur. It uses a number of sensors for touch and sound to detect interaction. The robot responds to the stimulation of interaction by making noises and moving.

3.2.2 Pleo dinosaur robot

The Pleo [1] is a commercially available pet dinosaur toy which was designed to have a lifelike appearance and adaptive behaviours. The 2008 model used in the experiment has a number of touch sensors on its head, chin, shoulders, back and feet, and audio and light sensors in its head. A range of actuators means it can respond to different types of interaction in different ways. The Pleo is covered with plastic which feels rubbery to touch.

3.2.3 Measures

All measures except the pen-and-paper evaluation form were administered via an online questionnaire on a tablet.

Quality of interaction with the other This was measured using items about how the participant felt during the interaction with the other person, and how the participant perceived the interaction itself:

Participants reported feelings experienced during the interaction by rating eight items from Leary, Kowalski, & Bergen [8] on a 7-point Likert scale from 1 (*not at all*) to 7 (*very much*). Factor analysis [5] reduced these items to two composite measures: *'relaxed', 'awkward', 'nervous'*, and *'confident'* loaded highly onto a factor of 'Confidence' during the interaction ($\alpha = .81$). *'Accepted', 'respected', 'disrespected'*, and *'rejected'* loaded onto a factor of ' Feeling Acceptance' during the interaction ($\alpha = .76$).

How the interaction was perceived was measured using 16 items adapted from Berry and Hansen[3], rated on a 7-point Likert scale from 1 (*not at all*) to 7 (*very much*). Factor analysis reduced these 16 items to four composite measures. First *'relaxed', 'smooth',* and *'natural'* loaded onto how 'Comfortable' the interaction felt ($\alpha = .84$). Secondly *'enjoyable', 'fun', 'pleasant', satisfying', 'intimate'*, and *'boring'* loaded onto a factor of the interaction 'Feeling Positive' ($\alpha = .86$). The third factor had loadings of *'upsetting', 'unpleasant'*, and *'annoying'* on a factor of the interaction 'Feeling Negative' ($\alpha = .65$). Finally *'forced', 'awkward', 'reserved'*, and *'strained'* loaded onto a factor of 'Difficulty' of the interaction ($\alpha = .86$).

Opinion of the other participant Participants answered the following questions adapted from Sprecher, Treger, Wondra, Hilaire, and Wallpe[16] about the interaction with the other participant and about the other participant on a 7-point Likert scale from 1 (*not at all*) to 7 (*very much*).

Liking of the other was measured with three items: '*How much did you like the other participant?*', '*How much would you like to interact with the other participant again?*', and '*How likeable did you find the other participant?*' ($\alpha = .86$)

Closeness to the other was measured with a single item: '*How close do you feel toward the other participant*?'

[5] Factor analysis for the purpose of dimension reduction was conducted using principal component analysis using oblimin rotation with each scale to create composite measures.

Perceived similarity was measured with two items: '*How much do you think you have in common with the other participant?*', and '*How similar do you think you and the other participant are likely to be?*'($\alpha = .86$)

Enjoyment of the interaction: This was measured with a single item:'*How much did you enjoy the interaction with the other participant*?'

Evaluation form The evaluation form consisted of a 10-item questionnaire about the robot which participants completed as a dyad. Five of the items were from Shibata, Wada, Ikeda, and Šabanović[14] and asked participants to indicate on a 7-point Likert scale how much they felt the words *'friendly', 'lively', 'expressive', 'natural',* and *'relaxing'* applied to the robot. The other five items were adapted from Wada, Shibata, Musha, and Kimura [20] and asked participants to answer on 7-point Likert scales the questions '*How cute/ugly do you find the robot?*', '*How much do you like the robot?*', '*How fun/boring is interacting with the robot?*', '*How much more would you want to interact with the robot?*' and '*How much do you want to touch the robot?*'.

3.3 Recording and coding behaviour

The interaction between the participants and the robot was covertly recorded in the experiment room with two Replay digital action cameras. Observed behavioural data will not be reported in this paper but will be detailed elsewhere.

3.4 Procedure

All participants were told that the study aimed to investigate people's opinions of different types of interactive robots, and that they would be asked to interact with and evaluate a robot. Participants were tested in dyads by a female experimenter. On arrival each participant was taken to a separate location to read the information sheet and provide consent to participate. Participants were told that they would meet another participant with whom they would evaluate a robot.

Both participants were first asked to complete a questionnaire (data not included in the current study). At this point the dyad was randomly assigned into either the active Paro, inactive Paro or Pleo conditions. Once both participants had completed the questionnaire, they were introduced to each other (as'the other participant you'll be evaluating the robot with') and together given an explanation of the robot evaluation task they were to undertake.

Participants were told that there would be a robot on the table in the room and were asked to interact with the robot together, in any way they wanted to, but to keep the robot off the floor. In the inactive Paro condition, participants were told that the robot would remain off for the duration of the task and that they would have the opportunity to see it turned on at the end of the session during individual debriefings. All participants were then told that there was an evaluation form on the table and were asked to complete the form together. The participants were told that they would be left and given 10 minutes to complete the task, after which the experimenter would knock on the door to the room and enter to take them to finish the experiment. The experimenter then took them into the room and before leaving, told them they could take a seat at the table.

Participants were given 10 minutes, which would provide sufficient time to complete the task and enable them to interact together beyond the scope of the evaluation. After the 10 minutes the experimenter entered the room and told the participants that the evaluation task was over. The participants were then taken to separate locations to complete a questionnaire to measure the quality of the interaction with the other and their opinion of the other participant. Subsequently the participants were individually thanked, debriefed, and informed of the covert recording which took place before providing their consent for use of the video data. In the inactive Paro condition participants were finally offered the opportunity to have a short interaction with the active Paro.

4 RESULTS

In this paper we report the quantitative data from the post-interaction questionnaire.

Table 1. Multilevel model of robot condition on quality of initial interactions and liking of other. $(*)$ indicates significance ($p < 0.05$), $(+)$ indicates a trend ($p < 0.1$)

	b	SE_b	p	95% CI
Feelings during interaction				
Confidence				
Active Paro vs Inactive Paro	0.26	0.26	0.335	-0.28,0.80
Active Paro vs Pleo	0.33	0.28	0.237	-0.22,0.89
Pleo vs Inactive Paro	-0.07	0.28	0.807	-0.64,0.50
Accepted				
Active Paro vs Inactive Paro	0.17	0.15	0.248	-0.12,0.47
Active Paro vs Pleo	0.18	0.15	0.247	-0.13,0.48
Pleo vs Inactive Paro	-0.01	0.15	0.970	-0.31,0.30
Perception of interaction				
Comfortable				
Active Paro vs Inactive Paro	0.16	0.29	0.585	-0.43,0.75
Active Paro vs Pleo	0.28	0.30	0.358	-0.33,0.89
Pleo vs Inactive Paro	-0.12	0.31	0.700	-0.74,0.50
Positive				
Active Paro vs Inactive Paro	0.46	0.23	0.049 $(*)$	0.00,0.92
Active Paro vs Pleo	0.42	0.24	0.083 $(+)$	-0.06,0.89
Pleo vs Inactive Paro	0.04	0.24	0.855	-0.44,0.53
Negative				
Active Paro vs Inactive Paro	-0.01	0.16	0.965	-0.33,0.31
Active Paro vs Pleo	-0.05	0.16	0.768	-0.38,0.28
Pleo vs Inactive Paro	0.04	0.17	0.804	-0.29,0.38
Difficult				
Active Paro vs Inactive Paro	-0.43	0.31	0.175	-1.05,0.20
Active Paro vs Pleo	-0.35	0.32	0.281	-0.99,0.29
Pleo vs Inactive Paro	-0.08	0.33	0.809	-0.73,0.58
Opinion of other				
Liking				
Active Paro vs Inactive Paro	0.33	0.22	0.135	-0.11,0.77
Active Paro vs Pleo	0.32	0.22	0.165	-0.13,0.76
Pleo vs Inactive Paro	0.01	0.23	0.948	-0.44,0.47
Closeness				
Active Paro vs Inactive Paro	-0.15	0.33	0.658	-0.81,0.52
Active Paro vs Pleo	0.36	0.34	0.297	-0.32,1.04
Pleo vs Inactive Paro	-0.51	0.35	0.150	-1.20,0.19
Similarity				
Active Paro vs Inactive Paro	0.00	0.31	0.992	-0.63,0.63
Active Paro vs Pleo	0.67	0.32	0.044 $(*)$	0.02,1.31
Pleo vs Inactive Paro	-0.66	0.33	0.049 $(*)$	-1.32,-0.00
Enjoyment of interacting				
Active Paro vs Inactive Paro	0.34	0.26	0.203	-0.19,0.86
Active Paro vs Pleo	0.60	0.67	0.031 $(*)$	0.61, 0.14
Pleo vs Inactive Paro	-0.26	0.27	0.350	-0.81,0.29

Dyadic analysis was required to account for the non-independence inherent in dyadic data [6]. This is due to the hierarchical structure of the data, with individuals nested into dyads. We used multilevel modelling in SPSS with the three robotic interaction conditions as predictors of the quality of interaction and liking of the other. The results are reported in table 1.

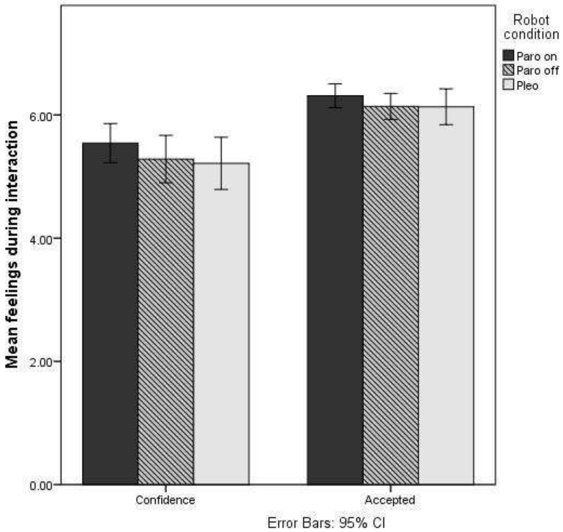


Figure 1. Feelings experienced by participants during the interaction for each robot condition

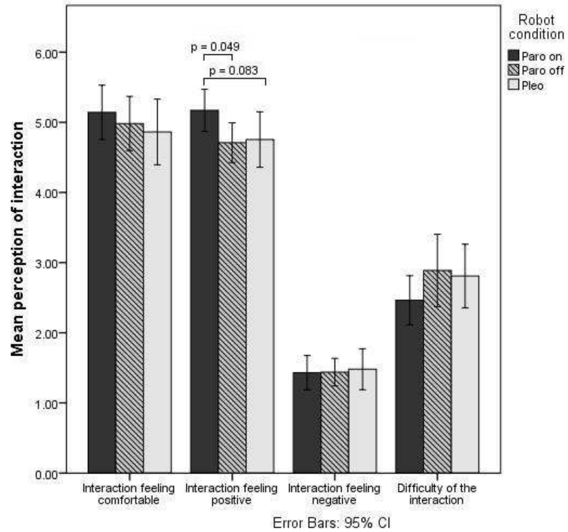


Figure 2. How the interaction was perceived for each robot condition

For the two factors measuring how participants felt during the interaction, no statistically significant differences between conditions were found, as seen in figure 1.

We found a significant difference between the active Paro and inactive Paro conditions for one quality of interaction factor, how positive the interaction felt. Participants in the active Paro condition had a significantly higher rating for positivity than those in the inactive Paro condition, $(b = 0.46, t(57.09) = 2.01, p = 0.049)$. In addition there was a positive trend toward significance for how positive the interaction felt for participants in the active Paro condition compared to those in the Pleo condition, $(b = 0.42, t(57.05) = 1.76, p = 0.083)$. There were no significant differences for how comfortable the interaction felt, how negative the interaction felt, and the difficulty of interaction. Figure 2 illustrates these results.

From the factors measuring participants' opinions of the other in Figure 3, perceived similarity to the other participant was significantly higher in the active Paro condition than in the Pleo condition $(b = 0.67, t(56.78) = 2.06, p = 0.044)$ but was significantly lower than the inactive Paro condition $(b = -0.66, t(56.16) = -2.01, p = 0.049)$. Participants in the active Paro condition had a significantly higher rating of enjoying interacting with the other than those in the Pleo condition, $(b = 0.60, t(56.89) = 2.21, p = 0.031)$

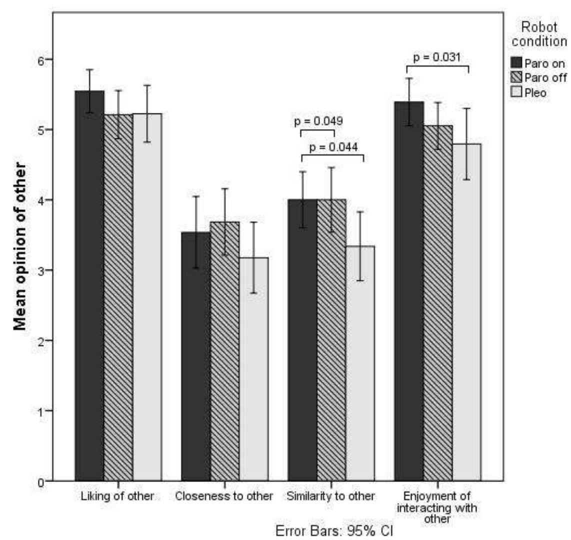


Figure 3. Participants' opinion of the other participant for each robot condition

5 DISCUSSION

The results from this study suggest that participants found the interaction with their partner more positive and had a higher opinion of their partner when interacting together with the active Paro, than with the inactive Paro or with the Pleo. This supports the hypotheses H1a and H1b.

However no results were found to support the hypotheses H2a or H2b, that participants who interact with the Pleo would have a stronger social mediation effect than the inactive Paro.

Of the hypotheses in H1, we found a significant result to partially support hypothesis 1a which concerns the quality of the interaction. The results show that participants who interacted with the active Paro had a greater generally positive feeling about the interaction with their partner than those who interacted with the inactive Paro. The trend between the active Paro and the Pleo, while still positive, was only near significant. A possible explanation for this is that when the Paro is active and interactive it is much more stimulating for both participants than when it was inactive, and provides a stronger focus for their interaction. The interactive Pleo may have been less effective due to the different appearance and texture, which is less cuddly and tactile and therefore less engaging.

Of the four factors to measure participants' opinions of the other two factors, similarity and enjoyment of interacting with the other person, show a significant effect. The significant effect was found between the active Paro condition and the inactive Paro and Pleo conditions which supports hypothesis 1b.

It is known that perceived similarity predicts interpersonal attraction [10], and has been found to predict long term attraction and the development of relationships in newly acquainted dyads[17]. Because interacting with the Paro, when active or inactive, has a larger impact on perceived similarity within pairs in this study, they may be judged as more likely to go on to form relationships than those with the Pleo. We suggest that this is because the Pleo has a more polarising effect than Paro, in which some people dislike it whereas others find it appealing, and is more likely to divide opinions during the interaction.

The higher ratings for the enjoyment of interacting with their partner for participants in the active Paro condition show that the experience of interacting together was improved by the presence of active Paro compared to the Pleo and inactive Paro.

In accordance with our primary hypothesis, these results show that the Paro, when active, is more effective as a social mediator and an ice-breaker for first-time interactions that the Pleo or inactive Paro. The lack of significant differences between the Pleo and inactive Paro conditions show that the second hypothesis is unsupported, and there is no difference between them as social mediators. This research suggests that the interactivity and the tactile texture are important factors of Paro which make it an and engaging and appealing object for individuals to interact over for the first time.

5.1 Limitations

A number of limitations need to be acknowledged in this study: the sample size did not provide the power to verify the findings with confidence. A number of results displayed the trend we hypothesised, and it is possible that larger numbers of participants would affect the significance values of these results.

The current study has only examined the social mediation effect of Paro with female participants and these results cannot be extended to male-male or female-male dyads. The response of males participants must be investigated as due to gender role norms, it is possible that males may respond more positively towards a robot which resembles a dinosaur to one resembling a seal.

One of the questions we posed was'Can the social mediation effect of the Paro be measured under laboratory conditions?' and these results show that some effect is measurable. However, while conducting the study under laboratory conditions allows a more controlled examination of the social mediation effect, the findings cannot be generalised to all social situations, and must be replicated in different situations to understand the possible applications of this effect.

Further work could include measures of personality and attachment in order to statistically control for individual differences in forming relationships. It would also be interesting to compare this study which used unacquainted dyads to one which uses people who already know each other.

6 CONCLUSIONS AND FURTHER WORK

The present study was designed to investigate the social mediation effect of Paro under controlled conditions. This research adds to the limited evidence which shows that robotic technologies can support social interaction between people. Our results suggest that when people interact together with Paro it helps provide a context in which to form a good first impression of their partner, and have a positive experience with them.

The findings of this study demonstrate that robotic technologies can support human-human interactions by encouraging social interaction and assist in the formation of relationships. More research is needed to fully understand this potential role for the further development of robot companions.

As the quantitative data in this study comes from self-report measures in the questionnaire, we expect the observed behavioural data from the covert video recording might highlight differences between interactions in robot conditions more clearly. The next stage of this study will be to examine the content of the interactions with the video data. Further research is needed to examine the social mediation effect of the Paro with its target users; older-aged adults, including those who are healthy and those with dementia. One application of the social mediation effect of Paro which has not been evaluated to date is its use in visits to care homes from family and friends. It would be valuable to investigate the role of Paro during these visits, and whether it leads to an increase in quality of the visitation time.

REFERENCES

[1] PLEO, 2010.

[2] R. Bemelmans, G. J. Gelderblom, P. Jonker, and L. De Witte. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. *Journal of the American Medical Directors Association*, 13(2):114 – 120.e1, Feb. 2012.

[3] D. S. Berry and J. S. Hansen. Personality, Nonverbal Behavior, and Interaction Quality in Female Dyads. *Personality and Social Psychology Bulletin*, 26(3):278–292, Mar. 2000.

[4] L. Giusti and P. Marti. Robots as social mediators: a study in the wild. *Gerontechnology*, 7(2), Apr. 2008.

[5] S. J. Hunt, L. A. Hart, and R. Gomulkiewicz. Role of Small Animals in Social Interactions between Strangers. *The Journal of Social Psychology*, 132(2):245–256, Apr. 1992.

[6] D. A. Kenny, D. A. Kashy, and W. L. Cook. *Dyadic Data Analysis*. 2006.

[7] C. D. Kidd, W. Taggart, and S. Turkle. A sociable robot to encourage social interaction among the elderly. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, number May, pages 3972–3976. IEEE, 2006.

[8] M. R. Leary, R. M. Kowalski, and D. J. Bergen. Interpersonal Information Acquisition and Confidence in First Encounters. *Personality and Social Psychology Bulletin*, 14(1):68–77, Mar. 1988.

[9] J. McNicholas and G. M. Collis. Dogs as catalysts for social interactions: Robustness of the effect. *British Journal of Psychology*, 91(1):61–70, Feb. 2000.

[10] R. M. Montoya, R. S. Horton, and J. Kirchner. *Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity*, volume 25. Dec. 2008.

[11] E. Mordoch, A. Osterreicher, L. Guse, K. Roger, and G. Thompson. Use of social commitment robots in the care of elderly people with dementia: A literature review. *Maturitas*, 74(1):14–20, 2013.

[12] H. Robinson, B. Macdonald, N. Kerse, and E. Broadbent. The psychosocial effects of a companion robot: a randomized controlled trial. *Journal of the American Medical Directors Association*, 14(9):661–7, Sept. 2013.

[13] A. Sharkey and N. Sharkey. Granny and the robots: ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14(1):27–40, July 2010.

[14] T. Shibata, K. Wada, Y. Ikeda, and S. Sabanovic. Cross-Cultural Studies on Subjective Evaluation of a Seal Robot. *Advanced Robotics*, 23(4):443–458, Jan. 2009.

[15] R. Sparrow and L. Sparrow. In the hands of machines? The future of aged care. *Minds and Machines*, 16(2):141–161, Aug. 2006.

[16] S. Sprecher, S. Treger, J. D. Wondra, N. Hilaire, and K. Wallpe. Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology*, 49(5):860–866, Sept. 2013.

[17] M. Sunnafrank and A. Ramirez. At First Sight: Persistent Relational Effects of Get-Acquainted Conversations. *Journal of Social and Personal Relationships*, 21(3):361–379, June 2004.

[18] S. Šabanović, C. C. Bennett, W.-L. Chang, and L. Huber. PARO robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. *IEEE ... International Conference on Rehabilitation Robotics : [proceedings]*, 2013:1–6, June 2013.

[19] K. Wada and T. Shibata. Robot therapy in a care house - Change of relationship among the residents and seal robot during a 2-month long study. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, volume 23 of *16th IEEE International Conference on Robot and Human Interactive Communication, RO-MAN*, pages 1250–1255, IEEE, Japan, Oct. 2007.

[20] K. Wada, T. Shibata, T. Musha, and S. Kimura. Robot therapy for elders affected by dementia. *IEEE Engineering in Medicine and Biology Magazine*, 27(4):53–60, July 2008.

[21] K. Wada, T. Shibata, T. Saito, and K. Tanie. Psychological and social effects of robot assisted activity to elderly people who stay at a health service facility for the aged. In *Proceedings - IEEE International Conference on Robotics and Automation*, volume 3 of *2003 IEEE International Conference on Robotics and Automation*, pages 3996–4001, Intelligent Systems Institute, AIST, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan, 2003.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Can Less be More? The Impact of Robot Social Behaviour on Human Learning

James Kennedy[1] and **Paul Baxter**[1] and **Tony Belpaeme**[1]

Abstract. In a large number of human-robot interaction (HRI) studies, the aim is often to improve the social behaviour of a robot in order to provide a better interaction experience. Increasingly, companion robots are not being used merely as interaction partners, but to also help achieve a goal. One such goal is education, which encompasses many other factors such as behaviour change and motivation. In this paper we question whether robot social behaviour helps or hinders in this context, and challenge an often underlying assumption that robot social behaviour and task outcomes are only positively related. Drawing on both human-human interaction and human-robot interaction studies we hypothesise a curvilinear relationship between social robot behaviour and human task performance in the short-term, highlighting a possible trade-off between social cues and learning. However, we posit that this relationship is likely to change over time, with longer interaction periods favouring more social robots.

1 INTRODUCTION

Social human-robot interaction (HRI) commonly focuses on the experience and perception of human users when interacting with robots, for example [2]. The aim is often to improve the quality of the social interaction which takes place between humans and robots. Companion robots increasingly aim not just to merely interact with humans, but to also achieve some goal. These goals can include, for example, imparting knowledge [11], eliciting behaviour change [17] or collaborating on a task [3, 13]. Studies with these goal-oriented aims often still apply the same principles for social behaviour as those without goals - that of maximising human interaction and positive perception towards the robot. The implicit assumption is often that if the interaction is improved, or the human perception of the robot is improved, then the chance of goal attainment will be increased as well.

In this paper, we focus on learning. In this context, we take learning to be the acquisition and retention of novel information, and its reuse in a new situation. This definition covers 3 areas from each of the 'Cognitive Process' (remember, understand, apply) and 'Knowledge' (factual, conceptual, procedural) dimensions of learning according to the revised version of Bloom's taxonomy [14]. Learning outcomes can depend on many different elements of behaviour, such as motivation [20] and engagement [4], which will also be considered here.

The remainder of this paper is structured as follows. First, studies in which social robots assist humans in learning will be reviewed, with the intention of showing the complex variety of results obtained when relating learning to the social behaviour of the robot (Section 2). Human-human interactions are then considered and are used as a basis to create a hypothesis about the relationship of robot social behaviour and human performance in tasks over both the long and short-term (Section 3). This leads to a discussion of the implications for HRI design in such contexts (Section 4).

2 MIXED LEARNING RESULTS IN HRI

One area of great potential in HRI is in using robots for education. However, mixed results are often found when using social robots to teach or tutor humans. Despite regular reports of liking robots more than virtual avatars, or preferring more socially contingent robots over those with less social capability, the human performance in learning tasks doesn't always reflect these positive perceptions [11, 12, 17, 22]. Conversely, significant cognitive gains have been found when comparing robots to virtual avatars, with varied amounts of contingent behaviour [15, 16]. Similar effects have been seen in compliance when comparing agents of differing embodiments [1]. Whilst the varied context and content to be learned between these studies could account for many of the differences in results, we suggest that the relationship between social behaviour and learning performance may be more complex than typically assumed.

Commonly, when behavioural manipulations are carried out on one or two cues, such as in a study by Szafir *et al.* varying the gestures and vocal volume that a robot uses, there are clear benefits to the human in terms of performance in learning tasks [26]. However, these positive benefits may be lost, or even reversed when larger manipulations to the social behaviour of the robot are applied, as in [12]. While it may be reasonably assumed that the effect of multiple individual cues is additive, this does not seem to be in accordance with the empirical evidence. Indeed, the proposition that social cues are perceived by humans as a single percept [29] considers individual social cues as providing the context for the interpretation of other social cues (recursively), leading to non-trivial interactions and consequences when multiple social cues are applied. There is thus the possibility that making large manipulations in social behaviour by varying multiple social cues simultaneously does not elicit the benefits that varying each of these cues individually would, as suggested by the data.

Human expectations of sociality will play a large role in an interaction with a robot. It has been suggested that a discrepancy between categorical expectations and perceptual stimuli could account for negative cognitive reactions [19]. We posit that humans don't necessarily expect to interact with a robot exhibiting social behaviours and that the discrepancy between their expectation and the reality of the interaction could create a cognitive reaction which impedes learning. This might explain some results showing a lack of improvement when social presence of an agent is increased (such as when going from a virtual avatar to a robot, as in [10, 17]), or when social behaviour

[1] Centre for Robotics and Neural Systems, Cognition Institute, Plymouth University, United Kingdom
email: {james.kennedy, paul.baxter, tony.belpaeme}@plymouth.ac.uk

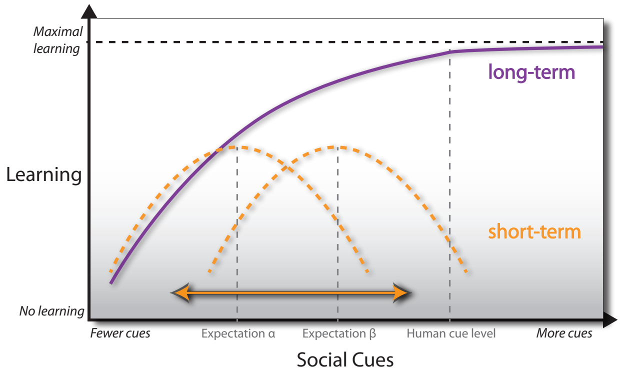


Figure 1: Hypothesised relationship between social behaviour (characterised by *immediacy* for example) as exhibited by a robot and its impact on the learning of a human in both the short and long-term. The position of the short-term curve is dependent on the humans' prior expectations of social behaviour (e.g. α is the expectation of fewer social cues from the robot than expectation β). Over time, these expectations normalise with reality, with increased use of social cues tending to lead to improved learning performance for the human interactant.

becomes more contingent, as in [12]. Expectation discrepancy would consequently lead to changes in the cognitive reaction over time as expectations change, and vary based on individuals, contexts, and so on; this is reflected in Figure 1 and will be expanded upon in Section 3.

Although there are many questions regarding learning in the context of HRI that remain unexplored, it would be useful to try and first create a testable hypothesis to attempt to explain why the results gathered so far are so varied. Whether this lies in social presence differences between virtual and physical robots, or in social behaviour manipulation between robot conditions, the main variable in all of the studies considered in this section is sociality. As such, we now consider how social behaviour might influence learning.

3 SOCIAL BEHAVIOUR AND LEARNING

In order to understand more about the nature of the relationship between social behaviour and learning, literature from human-human interaction (HHI) studies will now be introduced. Learning in the context of HHI has been under study for far longer than HRI, so longer-term research programmes have been carried out, and more data is consequently available.

When exploring the connection between learning and social behaviour in HHI literature, one behavioural measure repeatedly found to correlate with learning is 'immediacy'. Particularly applied to educational contexts, this concept has been long-established and validated across many cultures [18, 24] and age ranges [21]. Immediacy provides a single value definition of the social behaviour of a human in an interaction by characterising conduct in a range of verbal and non-verbal behavioural dimensions [23]. Immediacy could therefore prove a useful means of characterising robot social behaviour in HRI (as in [26]). Further, it has been shown that more immediate behaviours on the part of a human tutor increases cognitive learning gains [28]. However, the exact nature of the relationship between immediacy and cognitive learning gain is debated [5, 28].

Many HRI studies seem to implicitly assume a linear relationship between an increase in the number of social cues used or in social behaviour contingency and learning gains (or gains in related measures such as engagement, compliance, etc). Upon reviewing the literature concerning immediacy between humans, this has sometimes found to be the case [5], but more recent work has shown that this relationship may in fact be curvilinear [6]. A curvilinear relationship could go some way to explaining the mixed results found so far in HRI studies considering task performance with respect to robot social behaviour; it is possible that some studies make the behaviour *too social* and fall into an area of negative returns.

It is hypothesised that the curvilinear nature of immediacy may have been the effect observed in the study by Kennedy *et al.* in which a 'social' robot led to less learning than a robot which was actively breaking social expectations [12]. Over the short term, the novelty of social behaviour displayed by a robot may cause this kind of curvilinear relationship as has been observed in relation to immediacy [6]. As alluded to in Section 2, humans have a set of expectations for the sociality of the robot in an interaction. We would suggest that the greater the discrepancy between these expectations and the actual robot behaviour, the more detrimental the effect on learning. Individuals will have varied expectations, which is manifested in different short-term curves (Figure 1): the short-term curve shifts such that its apex (translating to the greatest possible amount of learning in the time-frame) is at the point where the expected and actual level of social cues is most closely matched. Prior interactions and the range of expectations created could also change the shape of the short-term curve, making the apex flatter or more pronounced depending on the variety of previous experiences.

However, when considering the interaction over the longer-term, such novelty effects wear off as the human adapts to the robot and their expectations change [7, 8, 25]. In this case we suggest that substantial learning gains could be made as the robot behaviour approaches a 'human' level of social cues; having attained a reasonable matching of expectation to reality, the robot can leverage the advantages that social behaviour confers in interactions, as previously suggested [9, 26]. Beyond this level, improvement would still be found by adding more cues, but the rate of increase is much smaller as the cues will require more conscious effort to learn and interpret. These concepts are visualised in the long-term curve seen in Figure 1.

4 PERSPECTIVES

So far, we have challenged the assumption that social behaviour has a simple linear relationship with learning by providing conflicting examples from HRI literature and also by tying concepts of social behaviour to the measure of immediacy from HHI literature. Given the regular use of HHI behaviour in generating HRI hypotheses, the non-linear relationship between immediacy and learning is used to hypothesise a non-linear relationship for HRI, particularly in the short-term (Figure 1).

A series of controlled studies would be needed to verify whether these hypothesised curves are correct. One particular challenge with this is the measuring of social behaviour. It is unclear what it is to be 'more' or 'less' social, and how this should be measured. This is where we propose that *immediacy* could be used as a reasonable approximation. All factors in immediacy are judgements of different aspects of social behaviour, which are combined to provide a single number representing the overall 'immediacy' (i.e. sociality of social behaviour) of the interactant. This makes the testing of such a hypothesis possible as the social behaviour then becomes a single dimension for consideration.

Of course, there are many other issues (such as robotic platform and age of human) which would need to be explored in this context, but with a single measure approximating sociality this would at least be possible. Providing an immediacy measure for robot behaviour makes it much easier to compare results between studies, allowing improved analysis of the impact of things such as task content and context, which are currently very difficult to disentangle when comparing results between studies. Literature from the field of Intelligent Tutoring Systems may be a useful starting point for future work to investigate specific aspects of learning activities due to their proven effectiveness across many contexts [27].

It should be noted that the aim of this paper is to highlight the potential directionality of the relationships involved between social cues and learning. There is not enough data available to represent the shape of the curves presented in Figure 1 with any great accuracy. The curves have been devised based on the few data points available from the literature, and following from concepts of immediacy and discrepancies of expectation, as explored in Sections 2 and 3.

5 CONCLUSION

We suggest that immediacy could be taken from the HHI literature to be validated and applied to HRI more extensively as it presents itself as an ideal means to facilitate comparison of highly varied social behaviour between studies. The large volume of immediacy literature in relation to learning and other contexts could also provide a firm theoretical basis for the generation and testing of hypotheses for HRI.

In this position paper we have shown through examples from HHI and HRI literature that the relationship between social behaviour and task outcome, specifically learning in the present work, for humans cannot be assumed to be linear. We hypothesise a model in which social behaviour not only has a non-linear relationship with learning, but also a relationship which changes over interaction time. Following the hypothesised model, we suggest that although in the short-term there may be some disadvantages for a robot to be maximally socially contingent, the benefits conferred by social behaviour as proposed by prior work will be seen in the long-term.

ACKNOWLEDGEMENTS

This work was partially funded by the School of Computing and Mathematics, Plymouth University, U.K., and the EU FP7 DREAM (FP7-ICT-611391) and ALIZ-E (FP7-ICT-248116) projects.

REFERENCES

[1] Wilma A Bainbridge, Justin Hart, Elizabeth S Kim, and Brian Scassellati, 'The effect of presence on human-robot interaction', in *Proc. RO-MAN'08*, pp. 701–706. IEEE, (2008).

[2] Mike Blow, Kerstin Dautenhahn, Andrew Appleby, Chrystopher L Nehaniv, and David C Lee, 'Perception of robot smiles and dimensions for human-robot interaction design', in *Proc. RO-MAN'06*, pp. 469–474. IEEE, (2006).

[3] Cynthia Breazeal, Nick DePalma, Jeff Orkin, Sonia Chernova, and Malte Jung, 'Crowdsourcing human-robot interaction: New methods and system evaluation in a public environment', *Journal of Human-Robot Interaction*, **2**(1), 82–111, (2013).

[4] Robert M Carini, George D Kuh, and Stephen P Klein, 'Student engagement and student learning: Testing the linkages*', *Research in Higher Education*, **47**(1), 1–32, (2006).

[5] Laura J Christensen and Kent E Menzel, 'The linear relationship between student reports of teacher immediacy behaviors and perceptions of state motivation, and of cognitive, affective, and behavioral learning', *Communication Education*, **47**(1), 82–90, (1998).

[6] Jamie Comstock, Elisa Rowell, and John Waite Bowers, 'Food for thought: Teacher nonverbal immediacy, student learning, and curvilinearity', *Communication Education*, **44**(3), 251–266, (1995).

[7] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C Schultz, et al., 'Designing robots for long-term social interaction', in *Proc. IROS'05*, pp. 1338–1343. IEEE, (2005).

[8] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro, 'Interactive robots as social partners and peer tutors for children: A field trial', *Human-Computer Interaction*, **19**(1), 61–84, (2004).

[9] A. Kendon, 'Some functions of gaze-direction in social interaction.', *Acta Psychol (Amst)*, **26**(1), 22–63, (1967).

[10] James Kennedy, Paul Baxter, and Tony Belpaeme, 'Children comply with a robot's indirect requests', in *Proc. HRI'14*, (2014).

[11] James Kennedy, Paul Baxter, and Tony Belpaeme, 'Comparing robot embodiments in a guided discovery learning interaction with children', *International Journal of Social Robotics*, (2014).

[12] James Kennedy, Paul Baxter, and Tony Belpaeme, 'The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning', in *Proc. HRI'15*, (2015).

[13] Hatice Kose-Bagci, Ester Ferrari, Kerstin Dautenhahn, Dag Sverre Syrdal, and Chrystopher L Nehaniv, 'Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot', *Advanced Robotics*, **23**(14), 1951–1996, (2009).

[14] David R Krathwohl, 'A revision of Bloom's taxonomy: An overview', *Theory into practice*, **41**(4), 212–218, (2002).

[15] Daniel Leyzberg, Sam Spaulding, and Brian Scassellati, 'Personalizing robot tutors to individual learning differences', in *Proc. HRI'14*, pp. 423–430, (2014).

[16] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati, 'The physical presence of a robot tutor increases cognitive learning gains', in *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, CogSci 2012, pp. 1882–1887, (2012).

[17] Rosemarijn Looije, Anna van der Zalm, Mark A Neerincx, and Robbert-Jan Beun, 'Help, I need some body: The effect of embodiment on playful learning', in *Proc. RO-MAN'12*, pp. 718–724. IEEE, (2012).

[18] James C McCroskey, Aino Sallinen, Joan M Fayer, Virginia P Richmond, and Robert A Barraclough, 'Nonverbal immediacy and cognitive learning: A cross-cultural investigation', *Communication Education*, **45**(3), 200–211, (1996).

[19] Roger K Moore, 'A Bayesian explanation of the 'uncanny valley' effect and related psychological phenomena', *Nature Scientific Reports*, **2**(864), (2012).

[20] Paul R Pintrich and Elisabeth van de Groot, 'Motivational and self-regulated learning components of classroom academic performance.', *Journal of educational psychology*, **82**(1), 33, (1990).

[21] Timothy G Plax, Patricia Kearney, James C McCroskey, and Virginia P Richmond, 'Power in the classroom VI: Verbal control strategies, nonverbal immediacy and affective learning', *Communication Education*, **35**(1), 43–55, (1986).

[22] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey, 'Comparing a computer agent with a humanoid robot', in *Proc. HRI'07*, pp. 145–152. IEEE, (2007).

[23] Virginia P Richmond, James C McCroskey, and Aaron D Johnson, 'Development of the Nonverbal Immediacy Scale (NIS): Measures of self- and other-perceived nonverbal immediacy', *Communication Quarterly*, **51**(4), 504–517, (2003).

[24] Vincent Santilli and Ann Neville Miller, 'The effects of gender and power distance on nonverbal immediacy in symmetrical and asymmetrical power conditions: A cross-cultural study of classrooms and friendships', *Journal of International and Intercultural Communication*, **4**(1), 3–22, (2011).

[25] Ja-Young Sung, Henrik I Christensen, and Rebecca E Grinter, 'Robots in the Wild: Understanding long-term use', in *Proc. HRI'09*, pp. 45–52. IEEE, (2009).

[26] Daniel Szafir and Bilge Mutlu, 'Pay attention!: Designing adaptive agents that monitor and improve user engagement', in *Proc. CHI'12*, pp. 11–20, New York, NY, USA, (2012). ACM.

[27] Kurt VanLehn, 'The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems', *Educational Psychologist*, **46**(4), 197–221, (2011).

[28] Paul L Witt, Lawrence R Wheeless, and Mike Allen, 'A meta-analytical review of the relationship between teacher immediacy and student learning', *Communication Monographs*, **71**(2), 184–207, (2004).

[29] Jamil Zaki, 'Cue integration a common framework for social cognition and physical perception', *Perspectives on Psychological Science*, **8**(3), 296–312, (2013).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Robots Guiding Small Groups: The Effect of Appearance Change on the User Experience

Michiel Joosse, Robin Knuppe, Geert Pingen, Rutger Varkevisser, Josip Vukoja, Manja Lohse and Vanessa Evers [1]

Abstract. In this paper we present an exploratory user study in which a robot guided small groups of two to three people. We manipulated the appearance of the robot in terms of the position of a tablet providing information (facing the group that was guided or the walking direction) and the type of information displayed (eyes or route information). Our results indicate that users preferred eyes on a display that faced the walking direction and route information on a display that faced them. The study gave us strong indication to believe that people are not in favor of eyes looking at them during the guiding.

1 Introduction

Social robots are designed to interact with humans in human environments in a socially meaningful way [3]. As a logical consequence, the design of robots often includes human-like features, e.g., heads or arms in order to generate social responses. It has been found that by using such anthropomorphic cues, people automatically have expectations of the robot's behavior [4].

However, the capabilities of robots differ from those of humans which allows them to use the anthropomorphic cues in different ways. For example, robot eyes can face the user while walking because the robot has other means (e.g., laser range finders) to detect the way to go. Thus, robots can walk backward. As eye contact has been shown to impact our image of others, and whether positive or negative, this being a sign of potential social interaction [6], robots facing users while guiding might actually be beneficial. On the other hand, literature indicates that people use a combination of head and eye movement to non-verbally indicate their direction [1] and users might expect robots to do the same.

Robots can also use non-anthropomorphic cues in different ways than humans, e.g. in the guiding context they can display route information rather than eyes. Related work found that visitors in historic places prefer a guide, as they would not have to worry about the route, or carry a map [2]. Therefore this could be beneficial for robots as well.

In the FP7-project SPENCER[2] we aim at developing a guide robot for a public place (airport) which will have a head and a screen. In this context, the questions arise which direction the head and screen should face when guiding a small group and what content should be displayed on the screen.

In related work, Shiomi et al. [5] conducted an experiment with the Robovie robot that drove either forward or backward while guiding participants in a mall (over a short distance). The overall finding in this experiment was that more bystanders joined when the robot moved backwards compared with frontwards, and that more people were inclined to follow the robot the entire time when moving backwards. In our work we are not so much interested in attracting people, but more in guiding people over a longer distance. Thus the question we pose here is how these design decisions impact the user experience in the process of guiding.

In this paper we present an exploratory study, in which we asked participants to follow a guide robot through a public lab space. This robot was equipped with a tablet (facing forwards, or facing the user) providing information to the participants. We were specifically interested in finding out which combination of tablet direction and type of information provided (eyes or route information) would yield the most positive user experience.

2 Method

In order to answer our research question, we designed an exploratory user study in which small groups of two to three participants were given a short guided tour by a robot.

2.1 Robot platform

For this study we attached a shell on top of a remote-controlled Robotino robot platform[3]. The height of the robot was 170cm and it drove at a speed of approximately 0.7 m/s. For purposes of this exploratory study, it was not deemed necessary to have the robot drive the path autonomously. Furthermore, the location of obstacles in the DesignLab changed from time to time (e.g. couches, chairs). As we were primarily interested in user experience ratings, the robot was remotely operated by an experimenter. Participants were not made aware of this before participating in the experiment.

2.2 Manipulations

We manipulated the direction of the tablet mounted on top of the robot and the information displayed on the tablet (Figure 1 and Table 1). In conditions A (Figure 1a) and B (Figure 1c) a set of blinking eyes was displayed on the tablet either facing the participants or the walking direction. In condition C we programmed the tablet to display route information, i.e., the remaining distance to the target (Figure 1e). A condition having the tablet mounted on the front of the robot, while displaying route information was deemed unnecessary as this would neither provide information for the participants following the robot, nor for other people present in the laboratory.

[1] Human Media Interaction group, University of Twente, the Netherlands, email: {r.a.knuppe, g.l.j.pingen, r.a.varkevisser, j.vukoja}@student.utwente.nl, {m.p.joosse, m.lohse, v.evers}@utwente.nl

[2] http://www.spencer.eu

[3] http://www.festo-didactic.com/int-en/learning-systems/education-and-research-robots-robotino/

(a) Condition A front (b) Condition A back (c) Condition B front (d) Condition B back (e) Condition C front (f) Condition C back

Figure 1: The appearance of the robot in the three conditions, showing the front and back side of the robot

Table 1: Overview of study conditions and number of participants

Condition	A	B	C
Tablet direction	Front	Back	Back
Tablet display	Eyes	Eyes	Time to destination
N	9	8	8
Group distribution	3x 3-person	1x 2-person 2x 3-person	1x 2-person 2x 3-person

2.3 Measures

In the post-experiment questionnaire user experience was assessed using a variety of measures.

All questions (except demographic- and open questions) were formulated as 5-point Likert-scaled items. General experience was assessed with eleven questions measuring among others if participants trusted that the robot knew where it was going, if it was clear where the robot was going and whether or not the robot was helpful in guiding someone. In this set of questions also the speed of the robot and volume of the audio messages were evaluated.

Five questions related to the physical appearance assessed the design, and specifically the height of the robot. Usability questions included questions related to users' expectancies of system capabilities and whether or not they were satisfied with the overall performance of the robot. Depending on the condition, this section included 5 (condition A), 6 (condition B), or 7 (condition C) questions.

Eight questions were included related to demographic information (age, gender, educational background) and familiarity with robots, social robots, and the premises where the test was conducted. A control question about the position of the tablet was included, and finally, we were interested in knowing whether or not the instructions provided were clear. Overall, this resulted in 30-32 questions

2.4 Procedure

Small groups of participants were recruited to participate in a guided tour of the DesignLab, a recently-opened lab of the University of Twente. Participants were given a briefing, after which they were given a tour of about five minutes through the lab. Participants were requested to follow the robot. No specific instructions were provided regarding the distance they should keep to the robot (Figure 4). The tour went past two points of interest (Figure 2, point B and C) where the robot provided a brief statement about the purpose using a text-to-speech engine. For example, when arriving at waypoint A, participants would see a tray with kinetic sand, and the robot would state that "The kinetic sand is made up of 98 percent sand, and 2 percent polyminethyl siloxane which gives it its elastic properties."

Afterwards the robot returned to the starting position where participants were requested to fill out the post-experiment questionnaire (Figure 2 point A). Following debriefing, participants were provided some candy as reward for their participation.

2.5 Participants

A total of 25 participants (14 males, 11 females) participated in the user study, with ages ranging from 17 to 40 (M=23.76, sd=5.93). All participants were students and staff from the University of Twente, primarily of Dutch (68%), German (8%) and Greek (8%) nationality. Participants had average experience with robots in general (M=2.84, sd=.90) and little experience with social robots (M=2.12, sd=1.09).

2.6 Data analysis

We calalculated means for all items. To compare between conditions, the data were first tested for normality. In case of normally distributed data, we report ANOVA's and T-tests in the results section, otherwise Kruskal-Wallis and post-hoc Mann-Whitney tests are reported.

3 Results

Overall, participants indicated they were quite satisfied with the robot: they believed the robot was helpful (M=4.47, sd=0.78), it

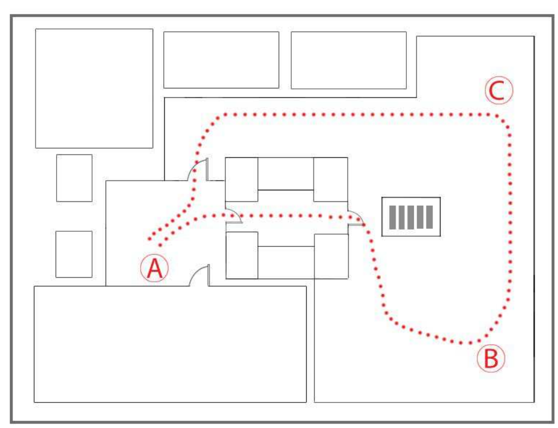


Figure 2: Layout of the laboratory showing start/end position (A) and two points of interest (B and C)

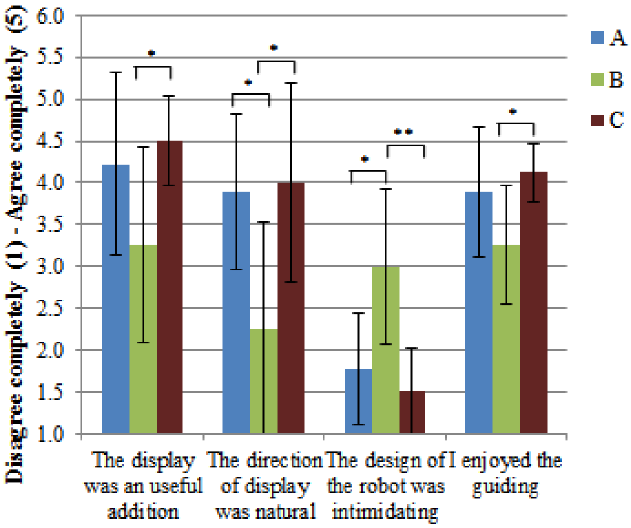


Figure 3: User experience ratings in the conditions; * indicates significance at the 0.05 level, ** at the 0.01 level

moved at a comfortable speed (M=3.12, sd=1.37), and participants trusted that the robot knew where it was going to (M=4.47, sd=0.78). These ratings did not differ significantly between conditions. Participants were moderately positive about the usability of the system: they felt comfortable using it (M=3.67, sd=1.05) and were satisfied by its performance (M=3.56, sd=0.77). No main effects or correlations were found including gender, age, robot experience and/or educational background.

Between conditions, Kruskal-Wallis tests indicated there were significant differences which were mostly due to the location of the tablet, thus between conditions A and C, versus condition B where the tablet was mounted on the front of the robot.

Post-hoc Mann-Whitney's indicated participants felt the direction of the screen was more appropriate in condition A (M=3.89, sd=.928) compared with B (M=2.25, sd=1.28), U=11.5; Z=-2.459, $p<0.05$. A similar effect was found between conditions B and C (M=4.0, sd=1.20), U=10.0, Z=-2.36, $p<0.05$ Furthermore, the design in condition B was more intimidating (M=3.00, sd=.97) compared with condition A (M=1.78, sd=.68), U=11.5, Z=-2.51, $p<0.05$ and condition C (M=1.50, sd=.54), U=6.00, Z=-2.885, $p<0.01$. Participants in condition C enjoyed the guiding more (M=4.13, sd=.35) compared with those in condition B (M=3.25, sd=.71), U=10.5, Z=-2.62, $p<0.05$.

With respect to the robot's appearance, participants felt that the body design matches the robot's function (M=2.71, sd=0.94). One of the interesting findings was that participants indicated the height was appropriate (M=4.21, sd=0.82). Informal sessions with participants indicated the robot would be too tall for a guiding robot, but in the end this was not the case. One of the reasons for this could be that participants' own average height was 177cm (sd=8.5cm), thus, most of them being taller than the robot.

4 Discussion & Conclusion

In this paper we presented an exploratory study into the effect of a robot's physical appearance on usability and user experience. Small groups of people were provided a short tour by a guide robot. Our results indicate that the location of the screen can be either forward

Figure 4: A small group of participants being guided by the robot

or backward, depending on the information displayed. In the case of eyes facing participants, our results showed that this was considered to be very unnatural and intimidating. On the other hand, when the tablet faced participants and route information was provided this was again evaluated as more useful. This might seem to be in contrast with the results of Shiomi et al. [5] who found that eyes facing participants are more effective to attract bystanders. However, we think this could be explained because in our setup the participants had already been introduced to the robot and asked to follow it.

Neither gender, age or experience with robots influenced the evaluation of the robots significantly, which could be due to small sample size.

Our future work will include a more interactive setup (e.g. provide participants some choices) during the tour. A second area of interest would be robot speed, and to investigate whether or not the speed of a guiding robot could be slower when guiding small groups compared with individual people. To conclude: the appearance of a guide robot can greatly influence user experience, something subtle as two eyes facing participants significantly decreases a robot's evaluation. Hence, more research is needed to even better understand how to design acceptable guide robots.

ACKNOWLEDGEMENTS

This research has been partly supported by the European Commission under contract number FP7-ICT-600877 (SPENCER).

REFERENCES

[1] Mark A Hollands, AE Patla, and JN Vickers, 'look where youre going!: gaze behaviour associated with maintaining and changing the direction of locomotion', *Experimental Brain Research*, **143**(2), 221–230, (2002).

[2] Daphne E Karreman, Elisabeth MAG van Dijk, and Vanessa Evers, 'Using the visitor experiences for mapping the possibilities of implementing a robotic guide in outdoor sites', in *RO-MAN, 2012 IEEE*, pp. 1059–1065. IEEE, (2012).

[3] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan, 'Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction', *Journal of communication*, **56**(4), 754–772, (2006).

[4] Manja Lohse, 'The role of expectations and situations in human-robot interaction', *New Frontiers in Human-Robot Interaction*, 35–56, (2011).

[5] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, 'A larger audience, please!: encouraging people to listen to a guide robot', in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction*, pp. 31–38. IEEE Press, (2010).

[6] Michael von Grünau and Christina Anston, 'The detection of gaze direction: A stare-in-the-crowd effect', *Perception*, **24**(11), 1297–1313, (1995).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

How can a tour guide robot's orientation influence visitors' orientation and formations?

Daphne E. Karreman[1], Geke D.S. Ludden[2], Elisabeth M.A.G. van Dijk[1], Vanessa Evers[1]

Abstract. In this paper, we describe a field study with a tour guide robot that guided visitors through a historical site. Our focus was to determine how a robot's orientation behaviour influenced visitors' orientation and the formations groups of visitors formed around the robot. During the study a remote-controlled robot gave short guided tours and explained some points of interest in the hall of Festivities in the Royal Alcázar in Seville (Spain). To get insight into visitors' reactions to the robot's non-verbal orientation behaviour, two orientations of the robot were tested; either the robot was oriented with its front towards the visitors, or the robot was oriented with its front towards the point of interest. From the study we learned that people reacted strongly to the orientation of the robot. We found that visitors tended to follow the robot tour guide from a greater distance (more than 3 meters away from the robot) more frequently when the robot was oriented towards the visitors than when it was oriented towards the point of interest. Further, when the robot was oriented towards the point of interest, people knew where to look and walked towards the robot more often. On the other hand, people also lost interest in the robot more often when it was oriented towards the point of interest. The analysis of visitors' orientation and formations led to design guidelines for effective robot guide behaviour.

1 INTRODUCTION

Several robots have been developed to give guided tours in a museum-like setting (some examples are described in [1]–[4]). These previously developed robotic tour guides did good jobs in their navigation and localization tasks, such as avoiding collisions with visitors or objects, and showing they were aware of the visitors' presence. While giving the tours, these robots captured the attention of visitors, had interactions with visitors and guided the visitors through smaller or larger parts of exhibitions. Studies reported some information about the visitors' reactions to the robot's actions which has led to knowledge on specific reactions of people to the modalities of these robots and behaviour shown by these robot designs.

Within the EU FP7 FROG project we were, among other innovations and application areas, interested in effective tour guide behaviour and personality for a robot guide. To find effective behaviours we started to examine the effect of single modalities on robot behaviour and visitor reactions to those behaviours. The question we wanted to answer with this study is: how does the robot orientation behaviour influence the orientations of the visitors, as well as the type of formations that (groups of) visitors form around the robot? The findings of the study we present in this paper led to guidelines to design behaviours (for FROG and other robots) that will influence visitors' reactions, such as orientation and group formations.

One way of creating robot behaviour is to copy human behaviour to a robot. A limitation of copying human tour guide behaviour to robots is that robots in general, and the FROG robot specifically, do not have the same modalities to perform actions that human tour guides perform. On the other hand, robots might have modalities to perform actions that human tour guides cannot perform. Therefore, we need to carefully study how and which robot modalities can effectively be used in interaction.

In previous studies, the reactions of the visitors were assumed to be similar to visitor reactions to human tour guides, but it turned out that these were different. For example, people often crowded around the robots [1], [2], [4], [5] or started to search for its boundaries by blocking the path [1] or pushing the emergency button [2], [6]. On the other hand, people often used their known human-human interaction rules to interact with the robots [2], even if the robots were not humanoid and people were informed that not all cues could be understood by the robot. Similar to robots that have been used in other studies, our FROG robot is not humanoid. We know that human tour guides influence visitor reactions of a group of visitors by using gaze behaviour and orientation [7]. Therefore, we are interested in visitors' reactions to a basic tour guide robot with limited interaction modalities. Also, we wanted to find out whether these reactions are similar to or different from visitor reactions to a human tour guide.

In this paper we will focus on the formation and orientation of visitors as a reaction to the robot orientation behaviour. We use the term formation to indicate the group structure, distance and orientation of the visitors who showed interest in the robot and/or the point of interest the robot described. In human guided tours, people generally stand in a common-focus gathering, a formation in which people give each other space to focus on the same point of interest, often a semi-circle [8]. For robot guided tours, we expected to find similar formations. However, from previous research we learned that single persons or pairs of visitors also joined the tour [9], [2]. Therefore, we considered the combination of distance and orientation of these individuals or pairs as formations as well. We assumed that people would be engaged with the robot or the explanation, when they were oriented towards the robot or the point of interest for a longer period of time. Hence, we also use the terms formation, orientation and engagement separately from each other in order to be specific in the description of the results.

[1] Human Media Interaction, Faculty of Electrical Engineering, Mathematics and Computing Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: {d.e.karreman, e.m.a.g.vandijk, v.evers}@utwente.nl
[2] Product Design, Faculty of Industrial Design Engineering, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands. Email: g.d.s.ludden@utwente.nl

In this paper, we first will discuss the related work on effects of robot body orientation, gaze behaviour and the use of several modalities in tour guide robots. Then we will present a field study where we aimed to find how robot orientation behaviour influences the group formations and orientations of the visitors. Next, we present will the results and discuss them. Finally, we will present design guidelines for non-verbal robot behaviour. The paper will end with a conclusion, in which we give directions for future research.

2 RELATED WORK

A tour guide robot for instance engages visitors and directs their attention to points of interest. This is similar to what human tour guides do intuitively. Human tour guides use their (body) orientation and give gaze cues to direct visitors' attention. However, most important are their subtle reactions to visitors' actions [7]. Kuzuoka et al. showed that a robot could effectively reshape the orientation of a visitor by changing its own orientation with a full body movement [10]. Also, human-like gaze cues can be successfully copied to robots, as shown by Yamazaki et al. They found that visitors showed higher engagement to a robot tour guide that used human-like gaze cues and its story than when the robot was not using these human-like gaze cues [11]. Sidner et al. found that head movements (and thus gaze cues) of the robot helped to keep people engaged during interaction [12]. Subtle gaze cues of robots can also be understood by people, as was shown by Mutlu et al. who let a robot describe an object among several other objects that were placed on a table. When the robot was "gazing" at the object it described, people found it easier to select the corresponding item [13].

The previously described body of work has focussed on copying two important types of cues that human guides use. However, robots are often able to apply a more diverse set of cues than body orientation and gaze cues alone. Different types of robots can use alternative modalities to give cues about their intentions. For example, if a robot uses a screen to convey information, visitors will stand close and orient themselves so that they can see the screen. However, when a robot uses arms to point and has no screen, visitors will probably orient themselves so they can easily see the robot and the exhibit the robot is pointing at.

Researchers have tried different modalities for museum robots to communicate intentions to their users. In the next paragraphs some examples of behaviour will be given to illustrate the effects of specific behaviours. The robot Rhino as developed by Burgard et al. blew a horn to ask visitors to get out of the way, which often had the opposite effect and made visitors stand in front of the robot until the horn sounded again [1]. Thrun et al. developed Minerva, the successor of Rhino. This robot did not have the problem that people clustered around when it wanted to pass, because it used several emotions and moods using its face and tone of voice. First, the robot asked in a happy and friendly state to get out of the way and if people did not react, the robot became angry after a while. With this behaviour Minerva was able to indicate its intentions and internal states successfully to the visitors [2]. However, the design of emotions and moods should be done carefully, as Nourbakhsh et al. found in the development of their robots. The robots Chips and Sweetlips showed moods based on their experiences that day. Visitors who only had a short interaction timeframe with the robots did not always understand these moods [4]. Touch screens and buttons have also been used for interaction purposes. These were found to make people stand closer to the robot, inviting them to interact with the buttons. This was for example found for the eleven Robox at the Expo.02 that were developed by Siegwart et al. [3]. However, buttons also can ruin the intended interaction. For example, Nourbakhsh et al. found that for the robot Sage [14] and Graf et al. for their robots in the museum of Kommunikation in Berlin (Germany) [6], people liked to push the emergency stop button and unintentionally stopped the robot from functioning.

All robots mentioned so far, had some interactive and social behaviour. However, specific guide behaviours - to engage multiple visitors and give information about exhibits - have still received little attention. To make a guided tour given by a robot a success, a smooth interaction between the robot guide and the visitors is essential, and therefore, interaction cues should be designed carefully.

Another challenge for museum robots is that they often have to interact with groups of people rather than with just one person. Research on group dynamics and behaviour of visitors gathering around a (dynamic) object in a museum setting or following a tour guide has revealed that visitors often stand in a specific formation (so-called F-formation) and react to each other and the (dynamic) exhibit (e.g. [7], [8], [15], [16]). For example, when a small group gathers around one person giving them information, they usually form a sort of (semi-) circle. In that way all group members can listen to the person who has the word [15]. Of course, the type of formation depends on the size of the group. However, the previously described formation is also recognizable when a human tour guide is guiding a (small) group of visitors and when people gather around a point of interest to all have the chance to see it [7]. When gathering around a museum object there are differences between gathering around interactive objects and static objects. When gathering around static objects, a lot of visitors get a chance to see the object at the same time. However, when gathering around interactive objects (often including a screen), fewer people can see the object at the same time [16], because people tend to stand closer to see the details shown on the screen or to directly interact with the (touch) screen. Museum exhibit designers tend to make the exhibits more interactive in order to keep the attention of the visitors, which also is effective for tour guide robots to attract visitors [4]. While these exhibits introduce more interactivity to the exhibition, it decreases the social interactions and collaborations between visitors [16]. Therefore, interactivity of robots should be designed for a larger group and other modalities than a screen/buttons should be used to shape the visitors' orientations and formations.

Our question is, can we design robots that have robot specific and intuitively understandable behaviour? To answer this question, robot designers have often resorted to directly copying human behaviour. In the design of other product categories, designers have often used anthropomorphism, (copying human forms and/or behaviour) in an abstract way rather than by directly copying. Subtly copying human forms or behaviour might likewise give cues about a product's intention and help people to understand the function of a product intuitively [17]. For robots, this implies that a robot does not have to directly resemble a human being, while it can still be capable of clearly communicating its intentions. Creating a robot with some anthropomorphic features does not necessarily mean that the robot needs to be human-like. However, to smooth the interaction human-like cues or features can be used in the design of robots [18]. Another question is, what should be designed first; the behaviour or the appearance of the robot. In most research on robots and their behaviour, the visual design for the robot was made first, and afterwards accompanying behaviour was designed. We decided to start from the other end. In this study, we used a very basic robot that showed some anthropomorphic behaviour in its body orientation. We were interested to find if and how people react to this behaviour while the appearance of the robot is far from human-like. In this way we expected to find some general guidelines for robot behaviour to influence people's reactions to the robot, while the options for the design of the robot are still multiple.

3 STUDY DESIGN

The goal of this study was to determine how orientation behaviour of a very basic robot influenced visitors' orientation and the formations groups of visitors formed around the robot. The orientation behaviour of the robot was manipulated, while other interaction features were limited. To evaluate how visitors reacted to the robot, we performed a study in the Royal Alcázar in Seville (Spain). The robot gave short tours with four stops in the Hall of Festivities of in the Royal Alcázar.

Participants

Participants of the study were visitors of the Royal Alcázar. At both entrances of the room, all visitors were informed with signs that a study was going on. By entering the room, visitors gave consent to participate in the study. It was up to them if they wanted to join the short tour given by the robot or not. Approximately 500 people (alone or in groups ranging from 2 to 7 visitors) interacted with the robot during the study.

Robot

The robot used for the field study was a four-wheeled data collection platform (see Figure 1). The body of the robot was covered with black fabric to hide the computers inside. A bumblebee stereo camera was visible at the top of the robot, as well as a Kinect below the bumblebee camera. The robot was remotely operated. The operator was present in the room, but he was not in the area where the robot gave tours. The robot was operated using a laptop. The laptop screen was used to check the status of the robot, while the keyboard was used to actually steer the robot. The interaction modalities of the robot were limited; the robot was able to drive through the hall, change its orientation and play pre-recorded utterances. The instruction "follow me" was visible on the front of the robot, and signs informing people about the research (in English and Spanish) were fixed to the sides of the robot.

The robot used for this study was very basic. We chose this particular robot to be able to determine the effects of body orientation on visitors' reactions without being influenced by other factors in robot design and behaviour (such as aesthetics of the robot, pointing mechanisms, visualisations on a (touch-) screen or active face modifications).

During the study we used a user-centred iterative design approach [19] for the behaviour of the robot. When the robot charged in between sessions, we discussed robot behaviours that had the intended effect and behaviours that did not work well. During the study we modified the explanation of the robot after session one, because it became clear that visitors did not understand where to look. A total of three iterations were performed. In all iterations only changes to the explanation of the robot were made, however the content about the points of interest remained the same.

Procedure

The tour given by the robot took about 3 minutes and 10 seconds. The points of interest chosen were all visible on the walls of the room (no exhibits were placed anywhere in the room), however the position of the points of interest on the walls differed in height. During a tour the distance to drive in between the points of interest also differed, from approximately two meters up to approximately five meters. This was done so we could see if there were different visitor behaviours when following the robot. However in this paper we will not focus on the results on following the robot.

When visitors entered in the Hall of Festivities, the robot stood at the starting place (1) (see Figure 2) and began the tour by welcoming the visitors and giving some general information about the room. When the robot finished this explanation, it drove to the next stop (about 3.5 meters away), asking the visitors to follow. At the next stop (2) the robot told the visitors about the design of the figures on the wall that were all made with tiles, after which it drove the short distance (about 2 meters) to the next exhibit. At the third stop (3) the robot told the visitors about the banner that hung high above an open door. At the end of this story the robot asked the visitors to follow after which it drove the long distance to the last stop (about 5 meters). Here (at point 4) it gave information about the faces visible on the tiles on the wall. Before ending the tour the robot drove back to the starting point (about 3.5 meters), informed the visitors the tour had finished and wished them a nice day.

After a while, when new visitors had entered the room, the robot started the tour again. During the study the robot tried to persuade visitors to follow it with the sentences "please follow me" and "don't be afraid", when visitors were hesitant. In all cases it was up to the visitors to decide whether they followed the robot or not. Visitors were never instructed to follow the robot by researchers present in the room.

As the study was performed in a real life setting, with uninformed naïve visitors, we sometimes had to deviate a bit from the procedure. The robot had defined places for stops. However, sometimes the robot had to stop close to the defined place, because people walked or stood in front of the robot.

Figure 1. Impression of the robot and visitors in the site

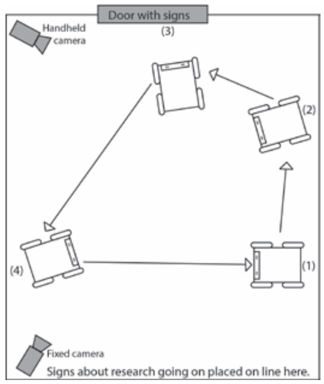


Figure 2. Layout of the tour

Another reason to deviate was when the robot lost the attention of all people who were following the tour. Then, it drove back to the starting place and started over again. If some visitors lost interest and left, but other visitors remained listening to the robot, it continued the tour.

When all visitors left the hall, or did not show any attention towards the robot, the trail was aborted, and restarted when new visitors entered the hall. Therefore the number of times the robot was presenting at each of the four exhibits was decreasing. The robot started the tour 87 times at the first exhibit, continued 70 times at the second exhibit. At the third exhibit the robot started its presentation 63 times and it finished the story only 58 times at the fourth exhibit. A total of 278 complete explanations at points of interest were performed (see table 1 for a specification of the actions per point of interest).

Manipulations

During the study, we manipulated the robot's orientation behaviour. Either the robot was orientated towards the point of interest or the robot was orientated towards the visitors. When it was orientated towards the point of interest, the front of the robot was in the direction of the point of interest. The points of interest were all located a few meters apart from each other. When the robot was oriented towards the visitors, its front was directed towards a single visitor or towards the middle of the group of visitors. See table 1 for a specification of the orientation of the robot per iteration and per point of interest.

In between the three iterations, some changes were made to the explanation by the robot. The explanations for the robot were developed in such a way that they could be used for both orientations of the robot. During the first iteration we observed that these explanation worked fine when the robot was oriented towards the points of interest. However, we found that it seemed unclear where to look when the robot was oriented towards the visitors. Therefore, for the second iteration, the explanations of the robot when oriented towards the visitors at points of interest two, three and four were modified. Information about where visitors had to look exactly to find the point of interest the robot explained about was added. As a result, the robot explained more clearly to the visitors "to look behind it" when it was orientated to the visitors and "to look here" when it was oriented towards the point of interest. Also, the sentences "please follow me" and "don't be afraid" were added to try to convince people to follow the robot to the next point.

Table 1: **Specification of manipulations**

	Robot actions	Point 1	Point 2	Point 3	Point 4
Iteration 1	**109**	**38**	**26**	**23**	**22**
To exhibit	*66*	*4*	*23*	*20*	*19*
To people	*27*	*27*	*0*	*0*	*0*
Excluded	*16*	*7*	*3*	*3*	*3*
Iteration 2	**90**	**25**	**24**	**22**	**19**
To exhibit	*42*	*0*	*10*	*17*	*15*
To people	*35*	*20*	*11*	*0*	*4*
Excluded	*13*	*5*	*3*	*5*	*0*
Iteration 3	**79**	**24**	**20**	**18**	**17**
To exhibit	*1*	*0*	*0*	*1*	*0*
To people	*65*	*16*	*18*	*16*	*15*
Excluded	*13*	*8*	*2*	*1*	*2*
Total	**278**	**87**	**70**	**63**	**58**
To exhibit	*109*	*4*	*33*	*38*	*34*
To people	*127*	*63*	*29*	*16*	*19*
Excluded	*42*	*20*	*8*	*9*	*5*

In the third iteration another modification was made to the explanation of the robot when it was oriented towards the visitors. The sentences were ordered in such a way that the robot would capture the attention of the visitors with something trivial, so people would not miss important parts of the explanations. All iterative sessions took about 1 hour and 40 minutes.

Data collection

During the study, the visitors were recorded with two cameras: a fixed camera that recorded the whole tour and a handheld camera that was used to record the facial expressions of the visitors close to the robot. Also, several visitors who followed (a part of) the tour were interviewed about their experiences. The interviews were sound recorded.

For this study only the data collected with the fixed camera was used, because the data from this camera gave a good overview of the room and the actions, orientation and formations of the visitors. We decided to not to use recordings from the cameras that were fixed on the robot, because their angle of view was limited to only the front of the robot. Using these recordings would not give us opportunities to study the behaviour of visitors who were next to or behind the robot (for example when the robot was oriented towards the exhibit), which in this study would lead to the loss of a lot of information on visitor orientation and formations. The proximity of the visitors was measured based on the number of tiles they stood away from the robot. Data collected through the short interviews was also not used in this analysis, because in this case we were only interested in how robot orientation influenced the actual orientation of visitors and their formations and less in their experience with the robot.

Data analysis

For the analysis, 236 robot actions of a total of 278 robot actions were used. Forty-two cases were excluded from analysis because no visitors were in the room or no robot was visible, because it was out of the angle of view of the camera, or the view was blocked by large numbers of visitors (for example a group with a human tour guide that did not show any interest in the robot). This resulted in 236 robot actions (278-42=236) in 3 iterations that were left for the analysis. The robot was oriented towards the exhibit while it explained 127 times, and the robot was oriented towards the visitors while it presented 109 times.

We were interested in the reactions of the visitors that might be influenced by the robot orientation during each if these 278 complete explanations at the points of interest. However, exact visitor behaviour to search for was not defined before the study. We performed a content analysis of the recordings from the fixed camera. We isolated robot actions -the moments that the robot stood close to a point of interest and presented about it- in the data for coding purposes. Coding of the data was done by using a Grounded Theory Method [20] and use of an affinity diagram [21] for the open coding stage. No exact codes were defined before the start of the analysis. We defined the codes based on the actions of the visitors found in the video recordings. Some examples of found codes are: "standing very close to the robot and oriented towards each other," "visitors standing in a semi-circle and robot oriented towards the exhibit," "visitors losing interest during the robot story and robot oriented towards the visitors," "visitors walking towards the robot and robot oriented towards exhibit." We used a count method to compare the reactions of the visitors during the robot actions between the different robot orientations and the different points of interest.

10 % of the data was double coded and we found an overall inter-rater reliability of a=0.662 (Cohen's Kappa), which indicates a substantial agreement between the coders. Hence, one coder finished the coding of the dataset that was used for analysis.

4 RESULTS

We found that visitors stood far away more often when the robot was oriented towards the visitors (31 times, 24.4% of all cases in this condition) than when the robot was oriented towards the point of interest (17 times, 15.6% of all cases in this condition). Further, no differences were found in formations of the visitors between both conditions. However, when the robot was oriented towards the visitors, just 18 times (14.2% of all cases in this condition) visitors walked towards the robot, while when the robot was oriented towards the point of interest visitors walked towards the robot 25 times (22.9 % of all cases in this condition). In both conditions and at all stops, a lot of people (78% of all cases) were just walking by, showing no attention for the robot at all. However, most of the time one or a few visitors had already joined the robot by then. A few times we observed that visitors waited until the robot was free again and then followed the tour. Also, when some of the visitors left the robot, others stayed to hear the rest of the explanation about the point of interest.

We found more differences between visitor formations when we focussed our analysis on the interactions in stops two, three and four, while excluding stop one. We decided to exclude stop one from our analysis, because at that stop the robot was always oriented towards the visitors and it was not explaining about a specific point in the room. We found that when the robot provided information about points of interest two, three and four, more people lost interest when the robot was oriented towards the point of interest (22 times, 21% of all cases in this condition) than when the robot was oriented towards the visitors (8 times, 12.5 % of all cases in this condition). Also, 6 times (9.4 % of all cases in this condition) visitors did not have a clue where to look when oriented towards the visitors. This was never the case (0% of all cases in this condition) when the robot was oriented towards the point of interest.

The number of visitors standing close to the robot was comparable between both conditions (5 times, 3.9% of all cases with orientation towards the visitors and 6 times, 5.5% of all cases with orientation towards the exhibit). However a difference was found between the exhibits. Only at stops one and two, did visitors stand really close to the robot when the robot was oriented towards the visitors. However, in the condition where the robot was oriented towards the point of interest people stood close to the robot at all stops. From reviewing the video, we observed that when people stood very close to the robot and the robot was oriented towards them, visitors only seemed to focus on the robot, while visitors focussed on the point of interest when the robot was oriented towards the point of interest.

Also we found some differences in visitor reactions between the different stops. Fewest visitors walked towards the robot at stop three (5 times; 9.3% of the cases in this condition), most did at stop four (16 times, 30.2% of the cases in this condition). Visitors lost interest in the story and the robot most often at stop three (14 times; 25.9% of all cases in this condition) and least often in stop four (6 times; 11.3% of all cases in this condition).

Looking only at the differences between the stops over both conditions, we found that many more single visitors and pairs joined the robot for at least one stop (86 times, 36.4% of all cases) than that people gathered around the robot in any group formation (38 times, 16.1% of all cases). We found that during 11 robot actions (4.7% of all cases) visitors stood less than 30 cm away from the robot. During 48 robot actions (20.3% of all cases) people stood more than 3 meters away from the robot. In 131 robot actions (55.5% of all cases) visitors stood between the 30 cm and 3 meters from the robot. Note that these cases can overlap, because there could be more than one visitor at the same time. In the rest of the cases no visitors or no robot were in the field of view or the visitors did not join the robot tour.

5 DISCUSSION

Influences of robot orientation

We found that visitors stood far away from the robot more often when the robot was oriented towards the visitors than when it was oriented towards the point of interest. Furthermore, we found that visitors tended to walk towards the robot more often when the robot was oriented towards the point of interest than when the robot was oriented towards the visitors. One possible explanation for this visitor reaction might be that visitors could not hear the robot well enough. However, we do not consider this a valid explanation in all cases, since people generally in both conditions followed the robot from a distance and they were able to hear the explanations of the robot. Therefore, we argue that it might be that the visitors felt that a distance was created by this specific orientation of the robot. This may have caused that people felt safer to approach the robot when it was oriented towards the point of interest. Perhaps, the robot kept people at a distance with its "eyes" when it was oriented towards the visitors. This finding is in line with findings from other studies that people walked closer to a robot that was not following them with gaze than when the robot was following them with gaze, as shown by Mumm and Mutlu [22]. Remarkable was that more people lost interest when the robot was oriented towards the point of interest than when the robot was oriented towards the visitors. As we argued before, the orientation of the robot towards the point of interest might have felt safer for people, at the same time, it might also have given them the feeling of being excluded, which made them leave the robot.

In stops one and two, several people were walking towards the robot, because the robot captured their attention and they were curious to see what it was for. Fewest visitors walked towards the robot at stop three, most did at stop four. Visitors probably did not have to walk to the robot in stop three, because it was really close to stop two. From stop three to stop four was the longest walk. Visitors who walked towards the robot in stop four were probably a bit reserved following the robot and therefore just walked to the robot when it had already started the next explanation. Apart from that, stop three was close to an open door, the entrance to the next room, therefore people who lost interest could easily walk away from the robot into the next room. When visitors followed to stop four, the last stop of the tour, they were likely to follow the robot the whole tour. We assume these visitors liked to hear the explanations of the robot and stayed with the robot until the final explanation, therefore fewer of them left the robot in stop four.

Visitor actions that were coded with "losing interest" showed that most of the time not all visitors lost their interest at the same moment. When one visitor of a pair or group walked away, the other(s) either followed the leaving person directly, stayed until the end of the explanation at that point or stayed until the end of the tour. This indicates that visitors of pairs or groups gave each other the time to do what they liked and that they did not have to leave together at the same moment. An advantage was that for most people it was clear that the robot just gave a short tour, so the people who left did not have to wait for a long time if the others stayed. In some cases we observed visitors discussing if they would follow the robot and in the end they decided that one would follow the tour, and that the other would wait outside the research area. It was important for the robot that when one visitor lost interest, most of the time the robot had other visitors (either close or far) who were still interested in the robot and the story, so it went on with the story.

We found a difference in the distance people kept from the robot and the orientation of the robot. Only at stops one and two, did visitors stand really close to the robot when the robot was oriented towards the visitors. However, when the robot was oriented towards the point of interest, visitors stood very close in all four stops. It seemed that when visitors stood very close to the robot and the robot was oriented towards them, visitors only had interest in the robot as an object and they tried to make contact with the robot (by waving at the robot or bringing their eyes on the same height as the lenses of the camera of the robot). We think this visitor behaviour mainly occurred at points one and two, because at these moments the robot captured people's attention. In stop three and four only visitors who were already following the tour seemed to be present and people who were only interested in the robot as an object did not disturb the robot guide and its visitors in these points. When visitors stood close and the robot was oriented towards the point of interest, the visitors probably could not hear the voice of the robot well enough to follow the story in the crowded area, while they were interested in the point of interest the robot presented about and wanted to hear the explanation.

Visitors who were interacting with the robot oriented towards them, sometimes appeared to have no clue where to look. This indicates that visitors were sensitive for the orientation of the robot. More verbal cues were added to the explanation of the robot in iterations 2 and 3. However, during these iterations, we still observed that when the robot was oriented towards them visitors got the clue where to look later than they expected. So, even though we changed the explanation of the robot to make more clear where to look and started with something trivial, just as human tour guides do [23], visitors did not readily understand where to look. This might be due to the length of the explanations of the robot. These were much shorter than explanations given by a human tour guide at a point of interest usually are. So, in general visitors had less time to focus again before they would miss something. The robot orientation towards the point of interest avoided this problem.

Visitor reactions to the "eyes" of the robot

Our observations showed that visitors were aware of the lenses of the camera on the robot and responded to them as if they were the eyes of the robot. This can for example be seen from the observation that some visitors waved at the camera when they arrived or when they left the robot. People also stood in front of the camera when they wanted to make contact with the robot. The observation that people are sensitive to the camera of a robot and orient in front of it was also made by Walters et al. [24]. These examples make clear that visitors react to the orientation of the robot and probably see the lenses of the camera as the eyes of the robot. Another observation that strengthens these conclusions is that visitors most often lost their interest in stop three. In this stop the explanation was difficult to understand because the story was about a banner that hung high in the room, above an open door. When the robot was oriented towards the exhibit, it seemed as if it was "looking" at a point in the other room because it was not able to tilt its orientation upwards. This confused the visitors, even when the robot was clear in its explanation about where to look.

Differences between robot guide and human tour guide

We found that visitors reacted differently to the robot tour guide than we would expect from observed reactions to a human tour guide. First of all fewer groups and more individual visitors or pairs of visitors joined the robot tour guide. Also, visitors seemed not prone to join strangers, but rather waited till the tour was finished and they could join a new tour.

Most visitors stood between 30 cm and 3 meters from the robot. When there were visitors standing very close or far away from the robot, there also could be visitors who stood at average distance (between 30 cm and 3 m) from the robot. While most visitors stood at an average distance, standing really close or staying at a distance differs from visitor behaviour shown when they follow a human tour guide. Most of the time visitors of a group of a human tour guide does not show that large difference in proxemics to a guide and often stand in a semi-circle to give everyone a chance to see the guide [7]. Also, Walters et al. [25] and Joosse et al. [26] showed in controlled experiments that people allowed different approach distances and appropriate proxemics for a robot than they allow for confederates. This leads to the conclusion that we cannot assume that people react the same to robot tour guides as to human tour guides.

Implications of study set-up

The study was performed in the wild which influenced the execution of the study and the manner of analysis. One disadvantage was that the situations of guiding could not be controlled. Also, less information of the visitors could be obtained. For example, we could not have extended questionnaires because people did not want to spend their time to filling these in.

We performed the study in several iterations in which we modified the explanation of the robot. Without these modifications to the explanations, we would not have been able to perform the manipulation of the orientation of the robot, because with the original explanation visitors did not seem to know where to find the point of interest when the robot was oriented towards them. This led to the following differences between the iterations. In iteration one the robot was mainly oriented towards the point of interest. In iteration two the modification of the explanation seemed insufficient, so the robot was mainly oriented towards the points of interest. In iteration three the robot was mainly oriented towards the visitors.

An advantage of the in-the-wild set-up of this study was that we observed the reactions of the visitors the way they would probably be if an autonomous tour guide robot were to be installed in the Royal Alcázar. The findings of this research were an important step for the development of FROG, because with in-the-lab studies with small groups of users, it would be difficult to create a similar environment including people who are acquaintances and strangers. Probably, we would also not have found how people react when the robot is already occupied by strangers, while in this set-up we did find interesting reactions of visitors in the real-world context.

Also, we used a very basic robot with limited interaction modalities. Nevertheless, the influence of body orientation and was largely observable in the visitor reactions. We expect that these factors will keep influencing visitor reactions when more robot modalities (such as arms to point, or a screen to show information) are added to the robot.

6 DESIGN IMPLICATIONS FOR ROBOT BEHAVIOUR

Findings described in the previous section led to the following set of design guidelines for the design of the non-verbal behaviour of a tour guide robot, that can be used irrespective of the visual design of the robot.

1) *Check for visitors standing far away when people close-by leave the robot during the explanation.*

The robot did not only catch the attention of people who were standing close such as we would expect with human tour guides. Visitors who chose to stay at a distance also followed the robot tour. Although these visitors were interested in the story and the robot, they did not want to be close. The tour guide robot should therefore not only focus on visitors nearby, but scan the surrounding once in a while and go on with the story or tour if it detects visitors who are not standing close, but show an orientation towards the robot and stay there during explanation. This behaviour of scanning the environment is even more important when visitors who are standing close all leave. Also, the robot should not rely solely on its detection of visitors by gaze (cameras directed to the front-side of the robot) to determine whether it should go on or stop the explanations, because in some situations the visitors tend to stand next to or behind the robot, while they are still interested in its story. The robot should be aware of these visitors and continue the explanation at the exhibit.

2) *Define behaviour of people standing close-by to decide whether to stop or to continue the story.*

For visitors who are standing close, the robot should make a distinction between people standing very close that are following the tour and people standing very close that show interest in the robot only. When people are still following the story, the robot should go on giving information. However, when people only show interest in the robot, the robot can decide to play with them a bit and show it is aware of the visitors being there. Possibly the robot can catch their attention for the story and change the playful or disturbing interaction to a guide-visitors interaction.

3) *Ask people to join the tour when they are hesitant to join strangers.*

The robot mainly attracted individuals and pairs who did not join other people who had started following the tour before them. People preferred to wait until others had left before they decided to join the tour. In other cases they just followed the tour from a distance, when other people were already close. This fits the purpose of the robot, however it would be nice if the small groups joined in order to all have an even better experience of the robot, because the robot cannot focus on all visitors close-by and far away. To do so, the robot can at certain moments in the story decide to scan for visitors and invite them to join.

4) *When camera lenses are clearly visible in the design of the robot, use them as eyes*

In our field study, a stereo bumblebee camera and a Kinect were clearly visible on the robot. Our experience in this study taught us that visitors see the stereo camera on top of the robot as the eyes of the robot. Therefore, when the camera cannot be hidden, the camera should be designed as eyes, including the design of gaze cues and gaze direction. Using these cues, especially when people expect them already, will probably smoothen the human-robot interaction. In our case, the FROG robot is not a humanoid robot, while the camera is visible. Therefore, we argue that a visible camera should be used as eyes of a robot, because this will support the mental model users will create of the robot.

7 CONCLUSION AND FUTURE WORK

To conclude, the orientation of the robot is important to shape the visitors' reactions. When it was clear to the visitors what to look at (mostly when the robot was oriented towards the exhibit), they became engaged more easily in the robot guided tour. However more people became interested in the robot when it was oriented towards the exhibit. Also, more people lost interest in the robot and the story when it was oriented towards the exhibit than when it was oriented towards the visitors. Therefore, keeping the attention should be done in a different way than capturing the attention of the visitors.

With this research we focused on visitors' orientation and group formations that visitors formed around the tour guide robot. However, in order to design robot behaviours for giving an effective tour, visitors' reactions when the robot is guiding them from one point of interest to the next should also be analysed, and guidelines about how to shape these should be developed. We will further use the recording from this study to analyse the visitor reactions to the robot guiding behaviour (e.g. following the robot from a distance or really close to the robot, hesitating to follow the robot) as well as visitor reaction at stops at points of interest while following the robot.

The present study has given us insight into how robot orientation and behaviour can influence people's formations and reactions. A future research question, is to find how the combined effects of robot behaviour and visual design of a robot will influence the number of people who stop to see the robot and eventually join the robot guided tour. In the future we will perform more elaborate evaluations including more robot modalities and behaviours.

ACKNOWLEDGEMENT

The research leading to these results received funding from the European Community's 7th Framework Programme under Grant agreement 288235 (http://www.frogrobot.eu/)."

We would like to thank N. Pérez-Higueras, R.R. Vigo and J. Pérez-Lara for preparing and controlling the robot. We would like to thank M.P. Joosse for his help with data analysis.

REFERENCES

[1] W. Burgard, A. B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun, "Experiences with an interactive museum tour-guide robot," *Artif. Intell.*, vol. 114, no. 1–2, pp. 3–55, 1999.

[2] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, F. Dellaert, D. Fox, D. Hahnel, C. Rosenberg, N. Roy, J. Schulte, and D. Schulz, "MINERVA: A second-generation museum tour-guide robot," in *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, 1999, vol. 3, no. May, pp. 1999–2005.

[3] R. Siegwart, K. O. Arras, S. Bouabdallah, D. Burnier, G. Froidevaux, X. Greppin, B. Jensen, A. Lorotte, L. Mayor, M. Meisser, R. Philippsen, R. Piguet, G. Ramel, G. Terrien, and N. Tomatis, "Robox at Expo.02: A large-scale installation of personal robots," *Rob. Auton. Syst.*, vol. 42, no. 3–4, pp. 203–222, Mar. 2003.

[4] I. R. Nourbakhsh, C. Kunz, and T. Willeke, "The Mobot Museum Robot Installations: A Five Year Experiment," in *2003 IEEE/RJS International Conference on Intelligent Robots and Systems*, 2003, pp. 3636–3641.

[5] M. Shiomi, T. Kanda, H. Ishiguro, and N. Hagita, "Interactive humanoid robots for a science museum," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, 2006, pp. 305–312.

[6] B. Graf and O. Barth, "Entertainment Robotics: Examples, Key Technologies and Perspectives," *Safety*, vol. 6, no. 6, pp. 8–12, 2002.

[7] K. Best, "Making museum tours better: understanding what a guided tour really is and what a tour guide really does," *Museum Manag. Curatorsh.*, vol. 27, no. 1, pp. 35–52, 2012.

[8] Adam Kendon, "Spacing and Orientation in Co-present Interaction," in *Development of Multimodal Interfaces: Active Listening and Synchrony; Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*, 2010, p. pp 1–15.

[9] B. Jensen, N. Tomatis, and L. Mayor, "Robots meet Humans-interaction in public spaces," *IEEE Trans. Ind. Electron.*, vol. 52, no. 6, pp. 1530–1546, 2005.

[10] H. Kuzuoka, Y. Suzuki, J. Yamashita, and K. Yamazaki, "Reconfiguring spatial formation arrangement by robot body orientation," in *Proceeding of the 5th ACM/IEEE international conference on Human-robot interaction - HRI '10*, 2010, pp. 285–292.

[11] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka, "Precision Timing in Human-Robot Interaction : Coordination of Head Movement and Utterance," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2008, pp. 131–139.

[12] C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh, "Where to Look: A Study of Human-Robot Engagement," in *Proceedings of the 9th international conference on Intelligent User Interfaces*, 2004, pp. 78–84.

[13] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita, "Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior," in *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction*, 2009, vol. 2, no. 1, pp. 69–76.

[14] I. R. Nourbakhsh, J. Bobenage, S. Grange, R. Lutz, R. Meyer, and A. Soto, "An affective mobile robot educator with a full-time job," *Artif. Intell.*, vol. 114, no. 1–2, pp. 95–124, Oct. 1999.

[15] P. Marshall, Y. Rogers, and N. Pantidi, "Using F-formations to analyse spatial patterns of interaction in physical environments," in *Proceedings of the ACM 2011 conference on Computer supported cooperative work - CSCW '11*, 2011, pp. 445–454.

[16] C. Heath, D. Vom Lehn, and J. Osborne, "Interaction and interactives: collaboration and participation with computer-based exhibits," *Public Underst. Sci.*, vol. 14, no. 1, pp. 91–101, 2005.

[17] C. DiSalvo and F. Gemperle, "From seduction to fulfillment: the use of anthropomorphic form in design," in *DPPI '03 Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, 2003, pp. 67–72.

[18] C. F. Disalvo, F. Gemperle, J. Forlizzi, S. Kiesler, and H. C. Interaction, "All Robots Are Not Created Equal : The Design and Perception of Humanoid Robot Heads," in *DIS '02 Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, 2002, pp. 321–326.

[19] J. D. Gould, T. J. Watson, and C. Lewis, "Designing for usability: key principles and what designers think," *Mag. Commun. ACM*, vol. 28, no. 3, pp. 300–311.

[20] J. M. Corbin and A. Strauss, "Grounded theory research: Procedures, canons, and evaluative criteria," *Qual. Sociol.*, vol. 13, no. 1, pp. 3–21, 1990.

[21] C. Courage and K. Baxter, *Understanding Your Users: A Practical Guide to User Requirements Methods, Tools, and Techniques*, vol. 2, no. 4. Morgan Kaufmann, 2005, p. 704.

[22] J. Mumm and B. Mutlu, "Human-Robot Proxemics : Physical and Psychological Distancing in Human-Robot Interaction," in *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on. IEEE*, 2011, pp. 331–338.

[23] D. E. Karreman, E. M. A. G. van Dijk, and V. Evers, "Contextual Analysis of Human Non-verbal Guide Behaviors to Inform the Development of FROG, the Fun Robotic Outdoor Guide," *Hum. Behav. Underst.*, vol. LNCS 7559, pp. 113–124, 2012.

[24] M. L. Walters, K. Dautenhahn, K. L. Koay, C. Kaouri, R. Te Boekhorst, C. Nehaniv, I. Werry, and D. Lee, "Close encounters: Spatial distances between people and a robot of mechanistic appearance," in *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots*, 2005, vol. 2005, pp. 450–455.

[25] M. L. Walters, K. Dautenhahn, S. Woods, K. L. Koay, R. Te Boekhorst, and D. Lee, "Exploratory studies on social spaces between humans and a mechanical-looking robot," *Conn. Sci.*, vol. 18, no. 4, pp. 429–439, 2006.

[26] M. Joosse, A. Sardar, and V. Evers, "BEHAVE: A Set of Measures to Assess Users' Attitudinal and Non-verbal Behavioral Responses to a Robot's Social Behaviors," *Soc. Robot.*, vol. LNAI 7072, pp. 84–94, 2011.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Performing Facial Expression Synthesis on Robot Faces: A Real-time Software System

Maryam Moosaei[1], Cory J. Hayes, and Laurel D. Riek

Abstract. The number of social robots used in the research community is increasing considerably. Despite the large body of literature on synthesizing facial expressions for synthetic faces, there is no general solution that is platform-independent. Subsequently, one cannot readily apply custom software created for a specific robot to other platforms. In this paper, we propose a general, automatic, real-time approach for facial expression synthesis, which will work across a wide range of synthetic faces. We implemented our work in ROS, and evaluated it on both a virtual face and 16-DOF physical robot. Our results suggest that our method can accurately map facial expressions from a performer to both simulated and robotic faces, and, once completed, will be readily implementable on the variety of robotic platforms that HRI researchers use.

1 Introduction

Robotics research is expanding into many different areas, particularly in the realm of human-robot collaboration (HRC). Ideally, we would like robots to be capable partners, able to perform tasks independently and effectively communicate their intentions toward us. A number of researchers have successfully designed robots in this space, including museum-based robots that can provide tours [10], nurse robots that can automatically record a patient's bio-signals and report the results [22], wait staff robots which can take orders and serve food [17], and toy robots which entertain and play games with children [30].

To facilitate HRC, it is vital that robots have the ability to convey their intention during interactions with people. In order for robots to appear more approachable and trustworthy, researchers must create robot behaviors that are easily decipherable by humans. These behaviors will help express a robot's intention, which will facilitate understanding of current robot actions or the prediction of actions a robot will perform in the immediate future. Additionally, allowing a person to understand and predict robot behavior will lead to more efficient interactions [18, 20].

Many HRI researchers have explored the domain of expressing robot intention by synthesizing robot behaviors that are human-like and therefore more readily understandable [29, 13, 21, 5]. For example, Takayama et al. [35] created a virtual PR2 robot and applied classic animation techniques that made character behavior more humanlike and readable. The virtual robot exhibited four types of behaviors: forethought and reaction, engagement, confidence, and timing. These behaviors were achieved solely by modifying the robot's body movement. Results from this study suggest that these changes in body movement can lead to more positive perceptions of the robot, such as it possessing greater intelligence, being more approachable, and being more trustworthy.

While robots like the PR2 are highly dexterous and can express intention through a wide range of body movements, one noticeable limitation is that there are some subtle cues they can not easily express without at least some facial features, such as confusion, frustration, boredom, and attention [16]. Indeed, the human face is a rich spontaneous channel for the communication of social and emotional displays, and serves an important role in human communication. Facial expressions can be used to enhance conversation, show empathy, and acknowledge the actions of others [7, 15]. They can be used to convey not only basic emotions such as happiness and fear, but also complex cognitive states, such as confusion, disorientation, and delirium, all of which are important to detect. Thus, robot behavior that includes at least some rudimentary, human-like facial expressions can enrich the interaction between humans and robots, and add to a robot's ability to convey intention.

HRI researchers have used a range of facially expressive robots in their work, such as the ones shown in Figure 1. These robots offer a great range in their expressivity , facial degrees-of-freedom (DOF), and aesthetic appearance. Because different robots have different hardware, it is challenging to develop transferable software for facial expression synthesis. Currently, one cannot reuse the code used to synthesize expressions on one robot's face on another [6]. Instead, researchers are developing their own software systems which are customized to their specific robot platforms, reinventing the wheel.

Another challenge in the community is many researchers need to hire animators to generate precise, naturalistic facial expressions for their robots. This is very expensive in terms of cost and time, and is rather inflexible for future research. A few researchers use commercial off-the-shelf systems for synthesizing expressions on their robots, but these are typically closed source and expensive as well.

Thus, there is a need in the community for an open-source software system that enables low-cost, naturalistic facial expression synthesis. Regardless of the number of a robot's facial DOFs, from EDDIE [34] to Geminoid F [9], the ability to easily and robustly synthesize facial expressions would be a boon to the research community. Researchers would be able to more easily implement facial expressions on a wide range of robot platforms, and focus more on exploring the nuances of expressive robots and their impact on interactions with humans and less on laborious animation practices or the use of expensive closed-source

[1] The authors are with the Computer Science and Engineering department, University of Notre Dame,{mmoosaei,chayes3,lriek}@nd.edu

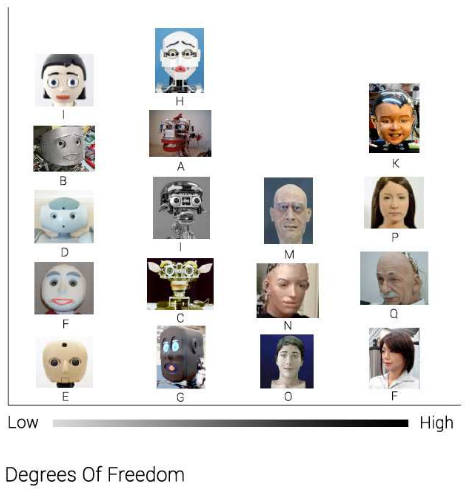


Figure 1. Examples of robots with facial expressivity used in HRI research with varying degrees of freedom. A: EDDIE, B:Sparky, C: Kismet, D: Nao , E: M3-Synchy, F: Bandit, G: BERT2, H: KOBIAN, F: Flobi, K: Diego, M: ROMAN, N: Eva, O: Jules, P: Geminoid F, Q: Albert HUBO , R: Repliee Q2

software.

In this paper, we describe a generalized software framework for facial expression synthesis. To aid the community, we have implemented our framework as a module in the Robot Operating System (ROS), and plan to release it as open source. Our synthesis method is based on performance-driven animation, which directly maps motions from video of a performers face onto a robotic (or virtual) face. However, in addition to enabling live puppeteering or "play-back", our system also provides a basis for more advanced synthesis methods, like shared gaussian process latent variable models [14] or interpolation techniques [23].

We describe our approach and its implementation in Section 2, and its validation in both simulation and on a multi-DOF robot in Section 3. Our results show that our framework is robust to be applied to multiple types of faces, and we discuss these findings for the community in Section 5.

2 Proposed method

Our model is described in detail in the following sections, but briefly our process was as follows: We designed an ROS module with five main nodes to perform performance driven facial expression synthesis for any physical or simulated robotic face.These nodes include:

S, a sensor, capable of sensing the performer's face (e.g., a camera)

P, a point streamer, which extracts some facial points from the sensed face

F, a feature processor, which extracts some features from the facial points coming from the point streamer

T, a translator which translates the extracted features from *F* to either the servo motor commands of physical platforms or the control points of a simulated head

$C: C_1 \dots C_n$, a control interface which can be either an interface to control the animation of a virtual face or motors on a robot.

These five nodes are the main nodes for our synthesis module. However, if desired, a new node can be added to generate any new functionality.

Figure 2 gives an overview of our proposed method. Assume one has some kind of sensor, (*S*), which senses some information from a person's (*pr*) face. This information might consist of video frames, facial depth, or output of a marker/markerless tracker. *pr* can be either a live or recorded performer. In our general method, we are not concerned about identifying the expressions on the *pr*'s face. We are concerned about how to use the expressions to perform animation/synthesis on the given simulated/physical face. *S* senses *pr* and we aim to map the sensed facial expressions onto the robot's face.

Basically, we use a point streamer *P*, to publish information from a provided face. Any other ROS node can subscribe to the point streamer to synthesize expressions for an simulated/physical face. A feature processor *F*, subscribes to the information published by the point streamer and processes this information. *F* extracts useful features out of all of the facial information published by the point streamer. Then, a translator, *T*, translates extracted features to control points of a physical/simulated face. Finally, a control interface $C: C_1 \dots C_n$ moves the physical/simulated face to a position which matches *pr*'s face.

2.1 ROS implementation

Figure 2 depicts the required parts for our proposed method. The software in our module is responsible for three tasks: (1) Obtaining input, (2) Processing input, (3) Actuating motors/control points accordingly

These responsibilities are distributed over a range of hardware components; in this case, a webcam, an Arduino board, a servo shield, and servo motors.

A local computer performs all processing tasks and collects user input. The data is then passed to the control interface, $C: C_1 \dots C_n$ which can either move actuators on an physical robot or control points on a virtual face. While Figure 2 shows the most basic version of our system architecture, other functionality or services can be added as nodes. Below, we describe each of these nodes in detail as well as the ROS flow of our method.

S, the sensor node, is responsible for collecting and publishing the sensor's information. This node organizes the incoming information from the sensor and publishes its message to the topic `/input` over time. The datatype of the message that this node publishes depends on the sensor. For example, if the sensor is a camera, this node publishes all incoming camera images. Examples of possible sensors include a camera, a Kinect, or a motion capture system. This node can also publish information from prerecorded data, such as all frames of a pre-recorded video.

P, the point streamer node, subscribes to the topic `/input` and extracts some facial points from the messages it receives. This node extracts some facial points and publishes them to the topic `/points`.

F, the feature processor node, subscribes to the topic `/points`. Node *F* processes all the facial points published by *P*. *F* extracts useful features from these points that can be used to map the facial expressions of a person to the physi-

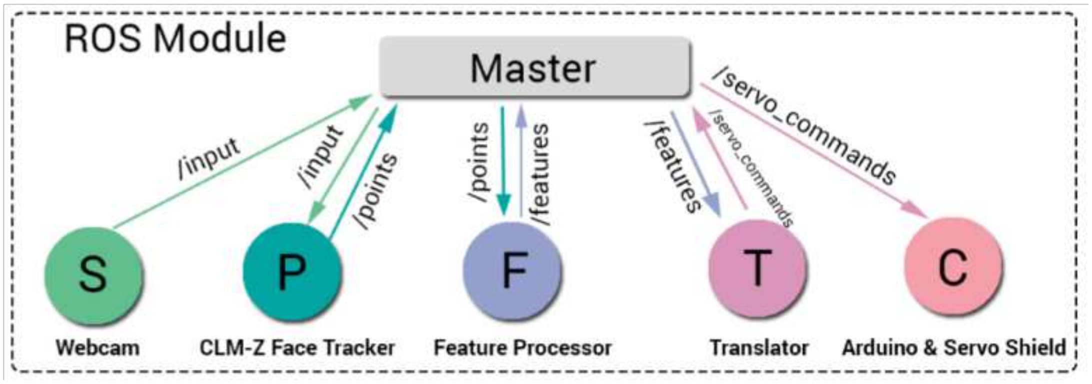


Figure 2. Overview of our proposed method.

cal/simulated face. This node publishes feature vectors to the topic `/features`.

T, the translator node, subscribes to the topic `/features`. and translates the features to DOFs available on the robot's face. Basically, this node processes the message received from the topic `/features` and produces corresponding movements for each of the control points on a robotic or virtual character face. This node publishes its output to the topic `/servo_commands`.

$C : C_1 \dots C_n$: The control interface node subscribes to the topic `/servo_commands` and actuates the motors of a physical robot or control points of a simulated face. We show $C : C_1 \dots C_n$ because a control interface might consist of different parts. For example, in case of a physical robotic head, the control interface might include a microcontroller, a servo shield, etc. We show the combination of all of these pieces as a single node because they cooperate together to actuate the motors. $C : C_1 \dots C_n$ subscribes to the topic `/servo_commands` which contains information about the exact movement for each of the control points of the robotic/simulated face. This node then makes a readable file for the robot containing the movement information and sends it to the robot.

2.2 An example of our method

There are various ways to implement our ROS module. In our implementation in ROS, we used a webcam as *S*. We chose the CLM face tracker as *P*. In *F*, we measured the movement of each of the facial points coming from the point streamer over the time. In *T*, we converted the features to servo commands for the physical robot and slider movements of the simulated head. In *C*, we used an Arduino Uno and a Renbotic Servo Shield Rev2 for sending commands to the physical head. For the simulated faces, *C* generates source files that the Source SDK was capable of processing.

We intended to use this implementation in two different scenarios: a physical robotic face as well as a simulated face. For a physical robot, we used our bespoke robotic head with 16 servo motors. For a simulated face, we used "Alyx", an avatar from video game Half-Life 2 from the Steam Source SDK. We describe each subsystem in detail in the following subsections.

2.2.1 *Point streamer P*

We employed a Constrained Local Model (CLM)-based face tracker as the point streamer in our example implementation. CLMs are person-independent techniques for facial feature tracking similar to Active Appearance Models (AAMs), with the exception that CLMs do not require manual labeling [12]. In our work, we used an open source implementation of CLM developed by Saragih et al. [1, 33, 11].

We ported the code to run within ROS. In our implementation, `ros_clm` (the point streamer) is an ROS implementation of the CLM algorithm for face detection. The point streamer `ros_clm` publishes one custom message to the topic `/points`. This message to the topic includes 2D coordinates of 68 facial points. This message is used to stream the CLM output data to anyone who subscribes to it.

As shown in the Figure 2, when the *S* node (webcam) receives a new image, it publishes a message containing the image data to the `/input` topic. The master node then takes the message and distributes it to the P node (ros_clm) because it is the only node that subscribes to the `/input` topic.

This initiates a callback in the *P* `ros_clm` node, causing it to begin processing the data which is basically tracking a mesh with 68 facial points over time. The `ros_clm` node sends its own message on the `/points` topic with the 2D coordinates of the 68 facial feature points.

2.2.2 *Feature processor F*

The *F* node subscribes to the topic `/points`. The *F* node receives these facial points. Using the position of two eye corners, *F* removes the effects of rotation, translation, and scaling. Next, in each frame, *F* measures the distance of each facial point to the tip of the nose as a reference point and saves 68 distances in a vector. The tip of the nose stays static in transition from one facial expression to the other. If the face has any in-plane translation or rotation, the distances of facial points from the tip of the nose will not be affected.

Therefore, any change in the distance of a facial point relative to the tip of the nose point over time would mean a facial expression is occurring. *F* publishes its calculated features to the topic `/features`.

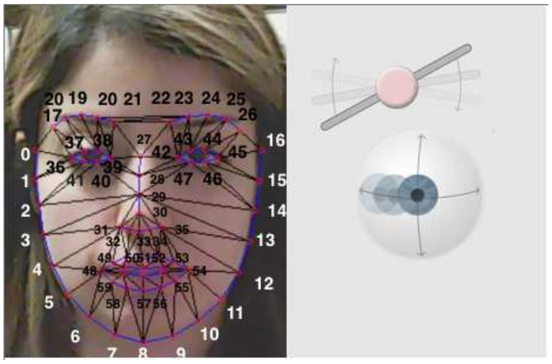


Figure 3. Left: the 68 facial feature points of CLM face tracker, Right: an example robotic eyebrow with one degree of freedom and an example robotic eyeball with two degrees of freedom

2.2.3 *Translator T*

The *T* node subscribes to the topic `/features` and produces appropriate commands for servos of a physical robot or control points of a simulated face. *F* keeps track of any changes in the distances of each facial point to the tip of the nose and publishes them to the `/features` topic. The *T* node has the responsibility of mapping these features to corresponding servo motors and servo ranges of a physical face, or to the control points of a simulated head. *T* performs this task in three steps. The general idea of these steps is similar to the steps Moussa et al. [28] used to map the MPEG-4 Facial Action Parameters of a virtual avatar to a physical robot.

In the first step, for each of the servo motors, we found a group of one or multiple CLM facial points whose movement significantly affected the motor in question. For example, as Figure 3 shows, the CLM tracker tracks five feature points on left eyebrow (22,23,24,25,26). However, the robot face shown in Figure 3 has only one motor for its left eyebrow. Therefore, the corresponding feature group for the robots left eyebrow, would be 22,23,24,25,26.

T converts the movement of each group of the CLM feature points to a command for the corresponding servo motor of a physical robot or control point of a simulated face. We used two examples in this paper, one with a simulated face and one with our bespoke robot. As an example, Table 1 shows the corresponding group of CLM points for each of the 16 servo motors of our bespoke robot

We averaged the movements of all of the points within a given group to compute only one number as the command for each motor/control point. To demonstrate this principle, our bespoke robot has a single motor for the right eyebrow. However, as Figure 3 shows, the CLM face tracker tracks five feature points on right eyebrow. If a performer raises their right eyebrow, the distance of these five points to the tip of the nose increases. We average the movements of these five points and use that value to determine the servo command for the the robot's right eyebrow.

Servo motors have a different range of values than that of feature points. Therefore, in the second step, we created a conversion between these values. The servos in our robot accept values between 1000 and 2000.

To find the minimum and maximum movement of each group of points associated with each servo, we asked a performer to make a wide range of extreme facial movements while seated in front of a webcam connected to a computer running CLM. For example, we asked the performer to raise their eyebrows to their extremities, or open their mouth to its maximum. Then, we manually modified the robot's face to match the extreme expressions on the subject's face and recorded the value of each motor. This way, we found the minimum and maximum movement for each group of facial feature points as well as for each servo motor.

In the last step, we mapped the minimum, maximum, and default values of the CLM facial points and the servo motors. Some servo motors had a reversed orientation with the facial points. For those servos, we flipped the minimum and maximum. In order to find values for a neutral face, we measured the distance of feature points to the tip of the nose while the subject had a neutral face. We also manually adjusted the robot's face to look neutral and recorded servo values.

Using the recorded maximum and minimum values, we applied linear mapping and interpolation (c.f., Moussa et al.) to find the criteria of mapping facial distances to servo values [28]. These criteria are used to translate facial points in each unseen incoming frame to the robot's servo values. The *T* node publishes a set of servo values to the topic `/servo_commands`.

2.2.4 *Control interface* $C : C_1 \ldots C_n$

The *C* node subscribes to the topic `/servo_commands` and sends the commands to the robot. The servo motors of our robot are controlled by an interface consisting of an Arduino UNO connected to a Renbotic Servo Shield Rev2. ROS has an interface that communicates with Arduino through the rosserial stack [2]. By using rosserial_arduino, a subpackage of rosserial, one can add libraries to the Arduino source code to integrate Arduino-based hardware in ROS. This allows communication and data exchange between the Arduino and ROS.

Our system architecture uses rosserial to publish messages containing servo motor commands to the Arduino in order to move the robot's motors. The control interface receives the desired positions for the servo motors at 24 frames-per-second (fps). For sending commands to the simulated face, C generates source files that the simulated face is capable of processing.

Table 1. The facial parts on the robot, and corresponding servo motors and CLM tracker points.

Facial Part	Servo Motor #	CLM Points
Right eyebrow	1	17,18,19,20
Left eyebrow	2	23,24,25,26
Middle eyebrow	3	21,22
Right eye	4 (x direction), 5 (y direction)	37,38,40,41
Left eye (x and y direction)	6 (x direction), 7 (y direction)	43,44,46,47
Right inner cheek	8	49,50
Left inner cheek	9	51,52
Right outer cheek	10	49,50,51
Left outer cheek	11	51,52,53
Jaw	12	56,57,58
Right lip corner	13	48
Left lip corner	14	54
Right lower lip	15	57,58
Left lower lip	16	55,56

3 Validation

To ensure our system is robust, we performed two evaluations. First, we validated our method using a simulated face (we used "Alyx", an avatar in the Steam Source SDK [3]). Then, we tested our system on a bespoke robot with 16 DOFs in its face.

3.1 Simulation-based evaluation

We conducted a perceptual experiment in simulation to validate our synthesis module This is a common method for evaluating synthesized facial expressions [9, 25]. Typically, participants observe synthesized expressions and then either answer questions about their quality or generate labels for them. By analyzing collected answers, researchers evaluate different aspects of the expressions of their robot or virtual avatar.

3.1.1 Method

In our perceptual study, we extracted three source videos of pain, anger, and disgust (total of nine videos) from the UNBC-McMaster Pain Archive [24] and MMI database [31], and mapped them to a virtual face. The UNBC-McMaster Pain Archive is a naturalistic database of 200 videos from 25 participants suffering from shoulder pain. The MMI database [31] is a database of images/videos of posed expressions from 19 participants who were instructed by a facial animation expert to express six basic emotions (surprise, fear, happiness, sadness, anger, and disgust). We selected pain, anger, and disgust as these three expressions are commonly conflated, and were replicating the approach taken by Riva et al. [32].

Using our synthesis module, we mapped these nine facial expressions to three virtual characters from the video game Half-Life 2 from Steam Source SDK. We used three different virtual avatars, and overall we created 27 stimuli videos 3 *(Expression: pain, anger, or disgust)* $\times$ 3 *(Gender: androgynous, male, and female)*. Figure 4 shows example frames of the created stimuli videos.

In order to validate people's ability to identify expressions synthesized using our performance-driven synthesis module, we conducted an online study with 50 participants on Amazon MTurk. Participant's ages ranged from 20-57 (mean age = 38.6 years). They were of mixed heritage, and had all lived in the United States for at least 17 years. Participants watched the stimuli videos in randomized order and were asked to label the avatar's expression in each of the 27 videos.

3.1.2 Results and discussion

We found that people were able to identify expressions when expressed by a simulated face using our performance-driven synthesis module (overall accuracy: 67.33%, 64.89%, and 29.56%[2] for pain, anger and disgust respectively) [19, 26]. Riva et al. [32] manually synthesized painful facial expressions on a virtual avatar with the help of facial animation experts, and found 60.4% as the overall pain labeling accuracy rate [32]. Although we did not set out to conduct a specific test to compare our findings to those of manual animation of the same expressions (c.f.

[2] Low disgust accuracies are not surprising; it is known to be a poorly distinguishable in the literature [8].

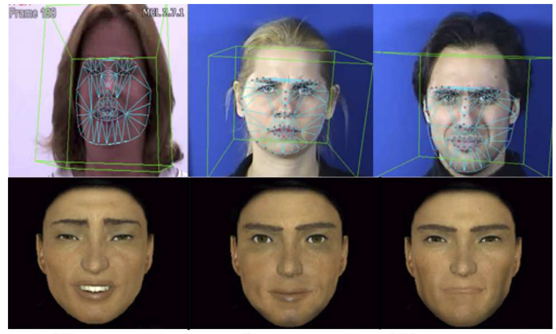


Figure 4. Sample frames from the stimuli videos and their corresponding source videos, with CLM meshes.

Riva et al. [32]), we found our synthesis method achieved arithmetically higher labeling accuracies for pain. These results are encouraging, and suggest that our synthesis module is effective in conveying naturalistic expressions. The next evaluation is to see how well it does on a robot.

3.2 Physical robot evaluation

To test our synthesis method with a physical robot, we used a 16-facial-DOF bespoke robot. Evaluating facial expressions on a physical robot is more challenging than on a simulated face because their physicality changes the physical generation of synthesis. Moving motors in real-time on a robot is far more complex a task due to the number of a robot's motors, their speed, and their range of motion.

We needed to understand if our robot's motors were moving in real time to their intended positions. Since the skin of our robot is still under development, we did not run a complete perceptual study similar to the one we ran in simulation. However, as we were testing how the control points on the robot's head moved in a side-by-side comparison to a person's face, we do not believe this was especially problematic for this evaluation.

3.2.1 Method

We ran a basic perceptual study with 12 participants to test both the real-time nature of the system, and the similarity between expressions of a performer and the robot. We recorded videos of a human performer and a robot mimicking the performer's face. The human performer could not see the robot. However, facial expressions made by the performer were transferred to the robot in real time.

The performer sat in front of a webcam connected to a computer. During the study, the performer was instructed to perform 10 face-section expressions, two times each (yielding a total of 20 videos). The computer instructed the performer to express each of the face-section expressions step by step. Face-section expressions were: *neutral*, *raise eyebrows*, *frown*, *look right*, *look left*, *look up*, *look down*, *raise cheeks*, *open mouth*, *smile*.

We recorded videos of both the performer and the robot mimicking the performer's face. Each video was between 3-5 seconds in length. We ran a basic perceptual study by using side-by-side

Table 2. Full results for each of the 10 face-section expressions.

Face-section expression	Average similarity score	s.d	Average synchrony score	s.d
Neutral	4.12	1.07	4.16	1
Raise eyebrows	4.33	0.86	4.25	1.13
Frown	4	1.02	4.08	1.24
Look right	4.54	0.5	4.66	0.74
Look left	4.5	0.77	4.37	1.08
Look up	2.83	1.29	3.7	1.31
Look down	3.54	1.41	4.45	0.77
Raise cheeks	3.79	1.4	4.25	0.85
Open mouth	4.12	1.16	4.62	0.56
Smile	2.79	1.14	4.41	0.91
Overall	3.85	1.28	4.3	1.01

comparison or "copy synthesis", which we have described in our previous work [27]. In a side-by-side comparison, one shows synthesized expressions on a simulated/physical face side-by-side with the performer's face to participants, and asks them to answer some questions [4, 36].

We showed side-by-side face-section videos of the performer and the robot to participants. Participants viewed the videos in a randomized order. We asked participants to rate the similarity to and synchrony with the performer's expressions and the robot expressions through use of a 5-point Discrete Visual Analogue Scale (DVAS). A five on the scale corresponded to " similar/synchronous" and a one to "not similar/synchronous" .

3.2.2 *Results and discussion*

Participants were all American and students at our university. Their ages ranged from 20-28 years old (mean age = 22 years). Eight female and four male students participated.

The overall average score for similarity between the robot and the performer expressions was 3.85 (s.d. = 1.28). The overall average score for synchrony between the robot and performer expressions was 4.30 (s.d. = 1.01).

Table 2 reports the full results for each of the 10 face-section expressions. The relatively high overall scores of similarity and synchrony between the performer and the robot expressions suggest that our method can accurately map facial expressions of a performer onto a robot in real-time. However, as this figure shows, we had a low average similarity score for *lookup* and *smile*.

One reason might be that the CLM tracker that we used in our experiment does not accurately track vertical movements of the eyes. Therefore, we could not accurately map the performer's vertical eye movements to the robot. Also, since our robot still does not have skin, its lips do not look very realistic. This might be a reason why participants did not find the robot's lip movements to be similar to the performer's lips movements.

4 General discussion

In this paper, we described a generalized solution for facial expression synthesis on robots, its implementation in ROS using performance-driven synthesis, and its successful evaluation with a perceptual study. Our method can be used both to map facial expressions from live performers to robots and virtual characters, as well as serve as a basis for more advanced animation techniques.

Our work is robust, not limited by or requiring a specific number of degrees of freedom. Using ROS as an abstraction of the code, other researchers may later upgrade the software and increase functionality by adding new nodes to our ROS module.

Our work is also a benefit to the robotics, HRI, and affective agents communities, as it does not require a FACS-trained expert or animator to synthesize facial expressions. This will reduce researchers' costs and save them significant amounts of time. We plan to release our ROS module to these communities within the next few months.

One limitation of our work was that we could not conduct a complete evaluation of our work on a physical robot, since its skin is still under development. Once the robot's skin is completed, we will run a full perceptual test. A second limitation was that the eye-tracking capabilities in CLM are poor, which may have caused the low similarity scores between the robot and performer. In the future as eye tracking technology advances (such as with novel, wearable cameras), we look forward to conducting our evaluation again.

Robots that can convey intentionality through facial expressions are desirable in HRI since these displays can lead to increased trust and more efficient interactions with users. Researchers have explored this domain of research, though in a somewhat fragmented way due to variations in robot platforms that require custom synthesis software. In this paper, we introduced a real-time platform-independent framework for synthesizing facial expressions on both virtual and physical faces. The best of our knowledge, this is the first attempt to develop an open-source generalized performance-driven facial expression synthesis system. We look forward to continuing work in this area.

5 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. IIS-1253935.

REFERENCES

[1] `github.com/kylemcdonald/FaceTracker`. Accessed: 2014-09-26.

[2] `http://wiki.ros.org/rosserial`. Accessed: 2014-09-26.

[3] Valve Software: Source SDK. `source.valvesoftware.com/sourcesdk.php`.

[4] B. Abboud and F. Davoine. Bilinear factorisation for facial expression analysis and synthesis. In *Proceedings of VISP '05*, pages 327–333, 2005.

[5] H. Admoni, A. Dragan, S. S. Srinivasa, and B. Scassellati. Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *Proceedings of HRI '14*, pages 49–56. ACM, 2014.

[6] A. Araújo, D. Portugal, M. S. Couceiro, and R. P. Rocha. Integrating arduino-based educational mobile robots in ros. In *Autonomous Robot Systems (Robotica) '13*, pages 1–6. IEEE, 2013.

[7] S. Baron-Cohen. Reading the mind in the face: A cross-cultural and developmental study. *Visual Cognition*, 3(1):39–60, 1996.

[8] D. Bazo, R. Vaidyanathan, A. Lentz, and C. Melhuish. Design and testing of a hybrid expressive face for a humanoid robot. In *IROS '10*, 2010.

[9] C. Becker-Asano and H. Ishiguro. Evaluating facial displays of emotion for the android robot geminoid f. In *WACI '11*, pages 1–8. IEEE, 2011.

[10] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke. Towards a humanoid museum guide robot that interacts with multiple persons. In *Humanoids '05*, pages 418–423. IEEE, 2005.

[11] S. W. Chew, P. Lucey, S. Lucey, J. Saragih, J. F. Cohn, and S. Sridharan. Person-independent facial expression detection using constrained local models. In *FG '11*, pages 915–920. IEEE, 2011.

[12] D. Cristinacce and T. F. Cootes. Feature detection and tracking with constrained local models. In *BMVC '06*, volume 2, page 6, 2006.

[13] P. F. Dominey and F. Warneken. The basis of shared intentions in human and robot cognition. *New Ideas in Psychology*, 29(3):260–274, 2011.

[14] C. H. Ek, P. Jaeckel, N. Campbell, N. D. Lawrence, and C. Melhuish. Shared gaussian process latent variable models for handling ambiguous facial expressions. In *American Institute of Physics Conference Series*, volume 1107, pages 147–153, 2009.

[15] P. Ekman. About brows: Emotional and conversational signals. In *Human ethology: Claims and limits of a new discipline: contributions to the Colloquium*, pages 169–248, 1979.

[16] R. El Kaliouby and P. Robinson. Generalization of a vision-based computational model of mind-reading. In *Affective computing and intelligent interaction*, pages 582–589. Springer, 2005.

[17] V. Emeli and H. Christensen. Enhancing the robot service experience through social media. In *RO-MAN '11*, pages 288–295. IEEE, 2011.

[18] F. Eyssel, D. Kuchenbrandt, and S. Bobinger. Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In *Proceedings of HRI '11*, pages 61–68. ACM, 2011.

[19] M. Gonzales, M. Moosaei, and L. Riek. A novel method for synthesizing naturalistic pain on virtual patients. *Simulation in Healthcare*, 2013.

[20] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. Chen, E. J. De Visser, and R. Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors*, 53(5):517–527, 2011.

[21] H. Knight. Eight lessons learned about non-verbal interactions through robot theater. In *Social Robotics*, pages 42–51. Springer, 2011.

[22] X. Li, B. MacDonald, and C. I. Watson. Expressive facial speech synthesis on a robotic platform. In *IROS '09*, pages 5009–5014. IEEE, 2009.

[23] J. Lieberman and C. Breazeal. Improvements on action parsing and action interpolation for learning through demonstration. In *Humanoids '04*, volume 1, pages 342–365. IEEE, 2004.

[24] P. Lucey, J. F. Cohn, K. M. Prkachin, P. E. Solomon, and I. Matthews. Painful data: The unbc-mcmaster shoulder pain expression archive database. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, pages 57–64. IEEE, 2011.

[25] D. Mazzei, N. Lazzeri, D. Hanson, and D. De Rossi. Hefes: An hybrid engine for facial expressions synthesis to control human-like androids and avatars. In *BioRob '12*. IEEE, 2012.

[26] M. Moosaei, M. J. Gonzales, and L. D. Riek. Naturalistic pain synthesis for virtual patients. In *Fourteenth International Conference on Intelligent Virtual Agents (IVA 2014)*.

[27] M. Moosaei and L. D. Riek. Evaluating facial expression synthesis on robots. In *HRI Workshop on Applications for Emotional Robots at the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2013.

[28] M. B. Moussa, Z. Kasap, N. Magnenat-Thalmann, and D. Hanson. Mpeg-4 fap animation applied to humanoid robot head. *Proceeding of Summer School Engage*, 2010.

[29] B. Mutlu, F. Yamaoka, T. Kanda, H. Ishiguro, and N. Hagita. Nonverbal leakage in robots: communication of intentions through seemingly unintentional behavior. In *Proceedings of HRI '09*, pages 69–76. ACM, 2009.

[30] K. Nagasaka, Y. Kuroki, S. Suzuki, Y. Itoh, and J. Yamaguchi. Integrated motion control for walking, jumping and running on a small bipedal entertainment robot. In *ICRA'04*, volume 4, pages 3189–3194. IEEE, 2004.

[31] M. Pantic, M. Valstar, R. Rademaker, and L. Maat. Web-based database for facial expression analysis. In *ICME '05*, 2005.

[32] P. Riva, S. Sacchi, L. Montali, and A. Frigerio. Gender effects in pain detection: Speed and accuracy in decoding female and male pain expressions. *EUR J PAIN*, 2011.

[33] J. M. Saragih, S. Lucey, and J. F. Cohn. Face alignment through subspace constrained mean-shifts. In *ICCV '09*. IEEE, 2009.

[34] S. Sosnowski, A. Bittermann, K. Kuhnlenz, and M. Buss. Design and evaluation of emotion-display eddie. In *IROS '06*, pages 3113–3118. IEEE, 2006.

[35] L. Takayama, D. Dooley, and W. Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of HRI '11*, pages 69–76. ACM, 2011.

[36] Q. Zhang, Z. Liu, B. Guo, and H. Shum. Geometry-driven photorealistic facial expression synthesis. In *Proceedings of SCA '03*, pages 177–186, 2003.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Gender, more so than Age, Modulates Positive Perceptions of Language-Based Human-Robot Interactions

Megan Strait and **Priscilla Briggs** and **Matthias Scheutz**[1]

Abstract. Prior work has shown that a robot which uses politeness modifiers in its speech is perceived more favorably by human interactants, as compared to a robot using more direct instructions. However, the findings to-date have been based soley on data aquired from the standard university pool, which may introduce biases into the results. Moreover, the work does not take into account the potential modulatory effects of a person's age and gender, despite the influence these factors exert on perceptions of both natural language interactions and social robots. Via a set of two experimental studies, the present work thus explores how prior findings translate, given a more diverse subject population recruited via Amazon's Mechanical Turk. The results indicate that previous implications regarding a robot's politeness hold even with the broader sampling. Further, they reveal several gender-based effects that warrant further attention.

Figure 1: Scenario: the humanoid MDS robot (Xitome Designs; left) instructs a confederate participant (right) on a brief drawing task.

1 INTRODUCTION

Natural language interactions with virtual and robotic agents are becoming increasingly pervasive, from virtual personal assistants (such as Apple's Siri agent), to socially assistive robots (e.g., elder care robots such as [4]). As the functionality of these artificial agents grows, so does the need to communicate with humans effectively to best serve the human interlocutor [12]. Surprisingly, however, there are very few attempts to date to carefully evaluate the different ways in which artificial agents could talk with humans in the context of a given task based on the agent's physical embodiment. For example, it is unclear whether an artifical agent, depending on its embodiment, should use imperatives when instructing humans (e.g., "turn right at the next intersection") or whether a more polite way of expressing an instruction is required (e.g., "we need to turn right at the next intersection"). Intuitively, a non-embodied agent like a navigation system might get away with syntactically simple, effecient imperatives, while a humanlike embodied robotic agent might have to employ more conventional forms of politeness.

Past work evaluating politeness in natural language interactions with robotic agents supports this intuition. Torrey and colleagues, for example, showed that the use of hedges (e.g., "*I guess*", "*probably*", and "*sort of*") and discourse markers – two "negative" politeness techniques – improves how people perceive a robot instructing a person via natural language. Specifically, they found that polite robots were viewed more positively than robots using more direct speech [22]. Even though negative politeness may be less noticeable than the *please*s of positive politeness, hedging indicates to the listener that the speaker is trying to mitigate the force of the request [7, 14].

[1] Human-Robot Interaction Laboratory, Tufts University, Medford, MA USA

Recent extensions of the above findings show that other negative politeness techniques (e.g., phrasing requests indirectly [9]), as well as positive (e.g., inclusive pronouns), suffice to improve perceptions of human-robot interactions (e.g., [6, 19, 21]). However, this research investigating human perceptions of robot politeness in human-robot interactions ([21, 22]) is predominately based on data drawn from the standard (and relatively homogeneous) university population.

Thus, whether and how these findings transfer to scenarios involving a population that is more diverse (e.g., economically, educationally), remains unknown. In particular, there are several factors (sociolinguistic, cultural, and demographic) in addition to politeness that have been found to modulate perceptions of natural language interactions (e.g., [3, 13, 16, 17, 20]). For instance, contrary to popular stereotypes, Japan is not as robot-positive as the US [2, 8].

Of particular relevance, is the growing amount of evidence that men (relative to women) hold significantly more positive towards robotic entities [5]. While both Torrey et al. ([22]) and Strait et al. ([21]) attempted to control for unintended effects due to gender, their participant samples were nevertheless imbalanced and thereby constrained in their ability to represent the general population. Hence, it is important to revisit these findings with explicit consideration of socio-demographic factors to understand what are their specific influences and how the findings extend beyond the university.

The goal of the present work was thus two-fold: (1) to investigate whether an extension of [21] with more diverse subject demographics would replicate the previously-observed effects of robot politeness (based on interaction observation), and further, (2) how the subject-based factors of age and gender specifically interact with those of the robot (e.g., the robot's use of polite communicatory cues).

To address these questions, we conducted a set of two online experiments via Amazon's Mechanical Turk with the aim of achieving greater diversity in people's age, and educational/geographical backgrounds, as well as more balanced gender demographics. In both, we presented videos depicting a robot instructing a person on a simple drawing task. We solicited people's reactions to these videos to determine the influence of a robot's politeness relative to any modulatory effects of a person's age and gender (**Experiment I**). Owing to a limitation of the first study, we conducted a follow-up to Experiment I to determine whether the findings hold given more naturalistic interaction settings (**Experiment II**).

	Comforting	*Considerate*	*Controlling*
Aggressive	−.15	−.11	**.68**
Annoying	**-.62**	−.27	.21
Comforting	**.73**	.30	−.13
Considerate	.21	**.63**	−.15
Controlling	−.11	−.16	**.52**
Eerie	**-.73**		.16
Likable	**.60**	**.59**	
Warm	.22	**.77**	−.24
Eigenvalues	3.63	1.16	.99
Variance Explained	.24	.44	.56

Table 1: Factor loadings for the three-factor EFA solution.

2 EXPERIMENT I

Based on the previous work outlined in the introduction ([21, 22]), we hypothesized that by using politeness modifiers in its speech, a robot would be perceived more favorably (as evidenced by higher ratings of *likeability* and reduced ratings of *aggression*) than a robot that uses more direct instructions. In addition, we generally explored the modulatory effects of a person's socio-demographic factors – in particular, age and gender – and how they interact with characteristics of a robot to influence perceptions of human-robot interactions.

To test our hypotheses and the age- and/or gender-based modulations thereof, we conducted a fully between-subjects investigation of the effects of a robot's communication strategy on observations of brief human-robot interactions – as influenced by a person's age and gender. In order to obtain a more diverse population than previously, we conducted our investigation online via Amazon's Mechanical Turk. Using a modification of the materials and methods developed in [21], we tasked participants with viewing a short video depicting a robot as it advised a person on creating a simple drawing. Following the video viewing, participants were prompted for their perceptions of the interaction, as rated on several dimensions regarding the likeability and aggression of the robot.

2.1 Materials & Methods

2.1.1 Participants & Procedure

839 participants were recruited via Amazon Mechanical Turk.[2] Prior to participating, subjects were informed the purpose of the study was to investigate factors that influence perceptions of human-robot interactions. Upon informed consent and subsequent completion of a demographic survey, the subject was shown one of 32 videos depicting a robot instructing a human confederate on a simple task. Following the viewing, the subject completed a 12-item questionnaire regarding his/her perceptions of the robot's appearance and behavior. Lastly, to assess attentiveness, participants completed a three-item check regarding salient details of the video clip.

Of these 839 participants, data from 329 were discarded due to several exclusion criteria: a restriction to limit participation to native english speakers (51 participants), and failure to complete the requested tasks (70) or failure on a three-item attention check (with a success threshold of 100%) to ensure participants viewed the presented video (208). Thus, our final sample included data from 510 participants (62% male) from 47 of 50 US states. The average age of this sample was 31.21 (SD=9.71), ranging from 18 to 76 years old. The most common level of education obtained was a bachelor's degree (45%), with an additional 36% of participants having some amount of college-level education. A small percent of participants reported having completed only high school (12%) and a smaller proportion reported obtaining more advanced degrees (7%). Participants also reported relatively high interest in robots (M=5.15, SD=1.32) – though low familiarity with robots (M=3.75, SD=1.49) – based on a 7-point Likert scale with $1{=}low$ and $7{=}high$.

[2] In anticipation of some loss in data due to exclusion criteria, we chose this sample size to achieve ≥ 15 useable observations in hypothesis testing.

2.1.2 Independent Variables

We employed a $2 \times 3 \times 2$ factorial design in which we systematically manipulated a robot's *politeness* in an advice-giving scenario, using the same conditions as those developed by Strait and colleagues ([21]). We also included participant *age* (three levels) and *gender* to investigate how they affect perceptions of the human-robot interaction. In total, we had the following three independent variables (IVs):

- **Politeness** of the robot's instructions (direct vs. polite). The *polite* condition entailed the robot giving instructions that contained one or more of both positive and negative politeness strategies, such as praise (e.g., "great job") and hedges (e.g., "a *kind of* large circle"). The *direct* speech condition employed the exact same instructions, but with the politeness modifiers removed.
- **Participant age** (three levels). We established three age categories based on a $1/3$ split of all the self-reported ages, resulting in a corresponding to the age of the standard university sample (M_1=22.81 years, SD=1.87), as well as two older adult categories (M_2=28.68, SD=1.99; M_3=42.16, SD=8.86).
- **Participant gender** (female vs. male).

2.1.3 Covariates

In addition to the above, we planned to carefully control for potential effects due to a person's motivations for completing the tasks (i.e., due to his/her purported *interest* in robots), as well as any effects due to characteristics of the stimulus set. To do so, we covaried three factors pertaining to the robot's physical embodiment:

- **Appearance** of the robot (two levels): the humanoid MDS (Xitome Designs) versus the less humanlike PR2 (Willow Garage).
- **Production modality** (synthetic vs. human speech), and
- **Gender** (female vs. male) of the robot's voice.

Thus, a total of four covariates – participants' **interest** in robots, the robot's **appearance** and the **gender** and **production modality** of the robot's voice, – were used in the analyses reported below.

2.1.4 Stimuli

A set of 32 videos (two conditions – polite versus direct speech – with 16 instances per condition) were constructed based on systematic manipulation of the robot-based IVs and covariates. Each video depicted a variant of a robot instructing a male human actor on a pen-and-paper drawing of a koala (cf. [21]). To avoid potential effects of affect, behavior, and/or movement (due to differences between the two robots' abilities), the robots were kept stationary. To avoid unintended effects due to a particular appearance, gender, voice, or the way in which the voice was produced, 16 video instances co-varying the robot's humanoid appearance (MDS versus the PR2), voice production modality (synthetic- versus human-produced speech) and voice gender (four voices – two female, two male) were created per condition. Four adult human actors comprised the set of human voices, with instructions to perform with flat affect. Synthetic voice production was performed using the native Mac OS X text-to-speech (TTS) software with four voices: "Alex","Ava", "Tom", and "Vicki". Following a between-subjects design, participants viewed only one video (selected randomly from the set of 32).

2.1.5 Dependent Variables

Of the set of 12 questionnaire items, three items – *task difficulty*, *interaction difficulty*, and *interest in interacting* – were considered as unique variables. On the remaining 9 items drawn from prior work (cf. [21, 22]), exploratory factor analysis produced a three-factor solution which showed a better fit ($\chi^2(7) = 13.36, p = .0638$) than a model where the variables correlate freely.

The criterion for retention of a questionnaire item was a factor loading of $\geq .50$ (see Table 1). We thus interpreted the three latent variables as the following: how **comforting** (four items – comforting, likable, -annoying, and -eerie; Cronbach's α=.83), **considerate** (three items – considerate, likable, and warm; α=.79), and **controlling** (two items – aggressive and controlling; α=.55) the robot was perceived. Items that were negatively correlated are indicated by $-$, and were automatically reversed in the computation of the latent constructs. Further, all dependent measures were normalized (to a scale between 0 and 1) prior to analysis.

2.2 Results

To assess the effects of the three IVs, between-subjects ANCOVAs were conducted on each of the dependent variables (taking into account the four covariates), with homogeneity of variance confirmed using Levene's test. All significant effects are reported below (with significance denoting $\alpha \leq .05$), and all post-hoc tests reflect a Bonferroni-Holm correction for multiple comparisons.

2.2.1 Comforting, Considerate, & Controlling

As expected, the *politeness* manipulation showed marginal ($p<.10$) to significant main effects on all three latent factors – *comforting*, *considerate*, and *controlling* (see Table 2, top). Similarly, participants' *gender* did as well (see Table 2, bottom); however, there were no significant main or interaction effects due to the participants' *age*.

Overall, both politeness and gender tended to increase ratings of the robot as *comforting* and *considerate*, and conversely, decrease those for *controlling*. However, these main effects were eclipsed by a *politeness* $\times$ *gender* interaction on both of the two positive factors: *comforting* ($F(1,498)$=4.57, p=.03, η^2=.01) and *considerate* ($F(1,498)$=6.97, p<.01, η^2=.01).

	DIRECT ($n = 254$)	**POLITE** ($n = 256$)	$F(1,498)$	p	η^2
Comforting	.13 (.37)	.19 (.38)	3.26	$= .07$	.01
Considerate	.46 (.16)	.54 (.17)	31.82	$< .01$	**.06**
Controlling	.25 (.17)	.20 (.16)	10.29	$< .01$	.02

	FEMALE ($n = 193$)	**MALE** ($n = 317$)	$F(1,498)$	p	η^2
Comforting	.21 (.40)	.11 (.36)	9.27	$< .01$	.02
Considerate	.53 (.16)	.48 (.16)	13.42	$< .01$	.03
Controlling	.20 (.16)	.26 (.17)	14.44	$< .01$	.03
Difficulty (t)	.17 (.16)	.21 (.18)	8.20	$< .01$	.02
Difficulty (i)	.24 (.23)	.28 (.21)	5.18	$= .02$	.01
Interest	.48 (.23)	.43 (.21)	4.74	$= .01$	.01

Table 2: Main effects of *politeness* (top) and *gender* (bottom), and relevant descriptive and inferential statistics.

In particular, the interactions showed that – while polite speech tended to improve participants' ratings – it did so primarily for women (see Figure 2, left and center). That is, a robot's use of polite speech significantly improved ratings of *comfort* when viewed by female observers (M=.29, SD=.39, n=94) relative to those by female observers of direct speech (M=.14, SD=.40, n=99; p=.04) and male observers of both direct (M=.11, SD=.34, n=155; p=.01) and polite speech (M=.11, SD=.38, n=162; p=.01). Similarly, though the polite robot significantly improved observers' ratings of *considerateness* for both female (M_{polite}=.59, SD=.16; M_{direct}=.47, SD=.17; p<.01) and male observers (M_{polite}=.50, SD=.17; M_{direct}=.45, SD=.15; p=.02), women's ratings were most improved relative to men's (p<.01).

With regard to perceptions of the robot as *controlling*, politeness was still broadly effective at decreasing ratings – regardless of the observer's gender, with polite robots receiving lower ratings relative to those more direct in their instructions (see Table 2, top). But, just being female helped as well: with women rating the robot as substantially less controlling than did men (see Table 2, bottom).

2.2.2 Difficulty & Interest

Gender further exerted significant main effects on the dependent variables regarding the perceived *difficulty* of both the task and interaction, as well as the observers' own *interest* in interacting with the depicted robot (see Table 2, bottom). In particular, female participants tended to rate both the task and interaction as less difficult than did males (see Table 2–bottom, *Difficulty*). Furthermore, they tended to show more interest in interacting with the robot agent than their male counterparts (see Table 2–bottom, *Interest*). There were no significant effects (main or interaction) due to *politeness* or *age*.

2.3 Discussion

Do people perceive a robot, which employs politeness modifiers in its speech, more favorably than one that uses more direct speech? Based on previous research by [21, 22], we expected that participants would rate a polite robot more favorably than one that is more direct in its instructions, as evidenced by higher ratings of positive constructs (e.g., *likability*) and lower ratings of negative constructs (e.g., *aggression*). Consistent with that work, the politeness manipulation here showed lower ratings of the robot as *controlling* and higher ratings of the robot as being *considerate* and *comforting*. In particular, our results

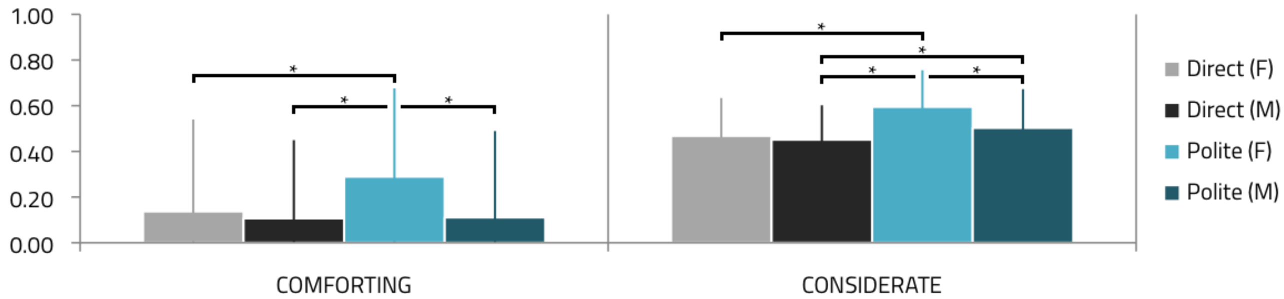


Figure 2: Interaction between robot *politeness* and participant *gender* on the three latent factors – the degree to which the robot was perceived as *comforting*, *considerate*, and *controlling*. Gray bars indicate the use of *direct* speech, versus blue, which indicates *polite* speech. Lighter bars indicate female participants (versus male participants, darker bars). All significant contrasts are shown (indicated by asterisks).

replicate and confirm those of prior work, even with a substantially more diverse subject population.

Does a person's age and gender further modulate perceptions of human-robot interactions? Based on previous suggestions that men and women view and respond to robots in significantly different ways [5], we evaluated the primary and modulatory effects of participants' age and gender. Participants' *gender* exerted a main effect on *all* dependent measures: how *comforting*, *considerate*, and *controlling* the robot was perceived as being, as well as how *difficult* both the task and interaction seemed and participants' *interest* in interacting with the depicted robot. In particular, female participants (relative to their male counterparts) showed more positive responding towards the robots and their interactions with the human confederate, as reflected by increased ratings of interest, comfort, and the robot's considerateness, as well as decreased ratings of the task/interaction difficulty and the robot's aggression. Further, interactions with the politeness manipulation showed that a robot's use of polite speech was effective at increasing women's positive attributions (the robot as being *comforting* and *considerate*), but not men's. Participant *age*, however, showed no main or interaction effects on *any* of the measures.

2.3.1 *Implications*

Prior work has suggested that a robot's use of politeness modifiers in its speech improves perceptions of human-robot interactions in advice-giving situations [21, 22]. Our results further replicate these findings (with respect to observation of human-robot interactions), and moreover, show the influence of politeness holds given a more general and representative population sample. In particular, our participants came from a wide variety of educational backgrounds (ranging from high school to advanced degrees) and geographical locations within the US (47 states).

In addition, we explicitely considered the effects of a person's age (ranging from the standard university age level to older adult) and their gender, to determine their influence and nature relative to the robot's politeness. This consideration of such socio-demographic factors revealed a methodological consideration for HRI studies – namely, that a person's *gender* should be taken into account when assessing perceptions of language-based human-robot interactions, as it is a modulating influence in addition to a robot's *politeness*.

This was expected, as previous research (e.g., [15, 18, 20]) has found that men exhibit more positivity towards robots than women. But, contrary to prior observations, our results indicate that women respond, in general, more positively towards the depicted robots. This may be due to the difference in the presentation the interactions as, here, video-recordings of human-robot interactions were evaluated by post-hoc observation, whereas, previous work has used scenarios involving the participatory and co-located interaction between the participant and robot of interest [16, 17, 20]. Alternatively (or in addition), it may be due to the difference in interaction: here, the robot interactants were depicted as instructing a human confederate; whereas, the human interactants in prior work were tasked with instructing or working with (rather than subservient to) the robot agent. Despite the conflicting differences in the nature of their effects, our findings add to the growing body of evidence implicating gender as an important methodological consideration in evaluating perceptions of human-robot interactions.

2.3.2 *Limitations & Future Directions*

Our approach to the investigation of perceptions of polite robots contributes a simple online task to assess the modulatory influences (or lack thereof) of a person's age and gender. In particular, the collection of data with broad socio-demographics augments in-laboratory studies that are limited to small, and relatively homogeneous, participant populations. This contribution here is significant because it replicates the previously reported influences of politeness, and further, sheds light on how such findings might transfer to the general population. That said, our approach also has several limitations (which underscore avenues for further research), three of which we discuss below.

Relevance. First, we note that the effect sizes for the given manipulations are relatively small. The magnitude of the effect of politeness on perceptions of the robot's considerate approaches a medium qualification (η=.10), but nevertheless, the implications of both robot politeness and participant gender are of limited weight. This may also suggest it is worth looking at the specific effects due to other factors such as a person's educational or geographic background (two socio-demographic items for which we did not control).

Mode of Evaluation. Another limitation of relevant consideration is how peoples' evaluations of the interactions were obtained. Here, the interactions were evaluated post-hoc by a third-party observer, who (by definition) was remotely located from the actual robot/interaction. This is particularly important to note, as it has been found that perceptions of human-robot interactions are further modulated by the interaction distance (remote versus co-located) and nature (observatory versus participatory) [21]. Thus, while the video-based interactions and online evaluations allowed us to sample from a broader demographic than that which is available locally, whether and how our gender-based findings apply to actual, co-located human-robot interactions warrants further investigation.

Stimuli. Lastly, there are a number of important limitations to the stimuli used and their presentation. Here the stimuli depicted brief (2 minute) interactions between an inanimate robot and a human confederate, which is an unrealistic interaction scenario in comparison to the intended usage of social robots.

In particular, prior work has shown that movement (however subtle) can impact the efficacy of interactions. For example, Andrist and colleagues have found that averting a robot's gaze (even for robot's without articulated eyes) can improve perceptions of the robot and their interactions [1]. Thus, with regard to the present study – though we limited movement to avoid unintended and/or differential influences (e.g., due to the robots' different capacities for actuation), the absence of movement itself might be affecting the current findings in unknown ways. For instance, the absence of attention-indicating gaze (e.g., looking at the participant when he/she is not performing a drawing instruction) might reduce positive attributions (e.g., considerateness) and/or increase negative attributions. This idea is supported by participants' open responses, which generally showed negative attitudes regarding the robot's lack of movement. Thus, there is the distict possibility that the lack of movement influenced perceptions in some way that may attenuate (or worse, decimate) other influences (e.g., due to politeness). With such considerations in mind, we moved to conduct a follow-up experiment to test the nature and magnitude of effects due to politeness and gender, when the robot was animated in a more naturalistic fashion.

3 EXPERIMENT II

Based on the considerations outlined in the previous section, we composed an exploratory follow-up investigation to Experiment I. We again conducted a between-subjects investigation of the effects of a robot's politeness (as influenced by a person's gender) on perceptions of human-robot interactions – but, with more naturalistic interactions. Specifically, we constructed a second set of video stimuli in which the robot was *animated* with attention-sharing and (human-like) idling movements, based on the naturalistic movements exhibited by a human instructing in such a context.

3.1 Materials & Methods

3.1.1 Participants & Procedure

437 additional participants were recruited via Amazon Mechanical Turk.[3] As in Experiment I, participants were told the purpose of the study was to investigate factors that influence perceptions of human-robot interactions. Upon informed consent and completion of a demographic questionnaire, the subject was shown one of 16 videos (similarly depicting a robot instructing a human confederate on a simple task). Following the viewing, the subject completed the 12-item questionnaire regarding his/her perceptions of the robot's appearance and behavior and the three-item check to assess whether the participant attended to the video.

Of these 437 participants, data from 176 participants were discarded due to: failure to complete the requested tasks (54) or failure on the attention check (122). Thus, our final sample included data from 261 participants (60% male) from 48 of the 50 US states. The average age of this sample was 32.45 (SD=10.45), ranging from 18 to 68 years old. The most common level of education obtained was similarly a bachelor's degree (44%), with an additional 37% of participants having some amount of college-level education. As in Experiment I, a small percent of participants reported having completed only high school (13%) and a smaller proportion reported obtaining more advanced degrees (6%). Participants again reported low familiarity (M=3.79, SD=1.49) with, but relatively high interest (M=5.33, SD=1.39) in robots.

3.1.2 Independent Variables

We again employed a fully factorial design, with the same independent variables as previously:

- The robot's **politeness** (direct vs. polite).
- **Participant age** (three levels): the standard university sample (M_1=22.85 years, SD=2.18), as well as two older adult categories (M_2=29.70, SD=2.01; M_3=43.98, SD=8.65).
- The participant's **gender** (female vs. male).

3.1.3 Covariates

We again planned to control for effects due to a person's *interest* in robots, as well as any due to characteristics of the stimulus set. As there was little variance explained by *production modality*, we excluded it from consideration to help reduce the overall number of videos to remake, thus reducing the number of observations needed to achieve similar sample sizes as Experiment I. As a result, we considered a total of three covariates in our analyses here: two factors pertaining to the robot's physical embodiment (the robot's *appearance* – MDS vs. PR2 – and *gender* of the robot's voice) and one factors pertaining to the participant (their *interest* in robots).

3.1.4 Stimuli

To increase the degree of observable presence/embodiment of the depicted robots, we recreated the videos from Experiment I[4] to animate the robots with select movements during the interaction. The movement modifications were intended to create a sense of "shared attention" and "idle" behaviors, based on the behaviors observed of a human instructor during pretesting of the drawing task with two people. In particular, the attentive behaviors were implemented such that the robot (MDS or PR2) moved its eyes (MDS) or head (PR2) up/down to focus on the human actor when giving instructions or on the actor's drawing (when the actor was drawing). Each robot also performed a set of idle behaviors (initiated based on random timers) throughout the interaction, based on their relative capacities for movement:

- *Blinking* (MDS only) – the MDS robot has two actuated eyelids that were closed and reopened ($500ms$) mimic human blinking.
- *Swaying* (MDS only) – the MDS has three degrees of freedom (DOF) on its center axis, allowing mimicry of slight head tilts (left/right and up/down positioning determined randomly at initiation of each tilt).
- *Breathing* (PR2 only) – the PR2, having fewer DOF with respect to its head movement, was limited to regular up/down undulation of its frontal laser. The rate of the laser movement approximated the average person's resting state heart rate ($70bpm$).

[3] In anticipation of data loss due to our exclusion criteria, we chose this sample size to again achieve ≥ 15 useable observations in hypothesis testing.

[4] As *production modality* was dropped from consideration, we recreated only a subset of the E1 videos – the 16 depicting a robot with a synthetic voice.

	DIRECT ($n = 130$)	POLITE ($n = 131$)	$F(1, 249)$	p	η^2
Comforting	.59 (.20)	.66 (.20)	6.72	= .01	.03
Considerate	.59 (.18)	.70 (.17)	26.27	< .01	**.11**
Controlling	.24 (.19)	.19 (.15)	6.31	= .01	.03

	FEMALE ($n = 104$)	MALE ($n = 157$)	$F(1, 249)$	p	η^2
Comforting	.68 (.20)	.58 (.20)	15.03	< .01	**.06**
Considerate	.68 (.18)	.62 (.18)	8.67	< .01	.03
Controlling	.20 (.16)	.24 (.17)	4.22	= .04	.02
Difficulty (t)	.24 (.23)	.29 (.23)	3.55	= .06	.01
Difficulty (i)	.26 (.28)	.35 (.27)	6.40	= .01	.03
Interest	.68 (.28)	.61 (.27)	3.45	= .06	.01

Table 3: Main effects of *politeness* (top) and *gender* (bottom), and relevant descriptive statistics, in Experiment II.

3.1.5 Dependent Variables

We used the same dependent measures as previously: task and interaction *difficulty* and *interest in interacting*, as well as how **comforting**, **considerate**, and **controlling** the robot was perceived as being.

3.2 Results

To assess the effects of robot *politeness* and participant *age*/*gender* – in the context of more naturalistic interactions – between-subjects ANCOVAs were conducted on each of the dependent variables (taking into account the four covariates), with homogeneity of variance confirmed using Levene's test. All significant effects are reported below (with significance denoting $\alpha \leq .05$), and all post-hoc tests reflect a Bonferroni-Holm correction for multiple comparisons.

3.2.1 Robot Politeness

As previously found, *politeness* exerted a significant effect on all three of *comforting*, *considerate*, and *controlling* DVs. Specifically, as expected based on Experiment I and previous literature, the robot's use of polite speech increased participants' comfort and their perceptions of the robot's considerateness. It also reduced perceptions of the robot as controlling (see Table 3, top).

3.2.2 Participant Age & Gender

Similarly, as Experiment I showed, *gender* improved perceptions along all dependent measures (see Table 3, bottom). Specifically, female participants continued here to (1) rate the robot as more *considerate* and less *controlling*, (2) indicate greater *comfort* and *interest in interacting* with the depicted robot, and (3) rate both the interaction and task as less difficult, than did their male counterparts.

Unlike the previous experiment, however, here participant *age* also showed a significant influence on *comfort* with the robot ($F(2, 249) = 3.19$, $p = .04$, $\eta^2 = .03$) and perception of it as *controlling* ($F(2, 249) = 4.07$, $p = .01$, $\eta^2 = .03$). Specifically, participants of the standard university age (young adults) indicated significantly less comfort with the robot ($M = .59$, $SD = .22$, $n = 87$) than the oldest participants ($M = .67$, $SD = .19$, $n = 92$; $p < .01$). Conversely, the younger participants also rated the robot as significantly more controlling ($M = .26$, $SD = .19$, $n = 87$; $p = .01$) than did either of the two older age groups – middle adults ($M = .19$, $SD = .16$, $n = 82$) and older adults ($M = .20$, $SD = .16$, $n = 92$).

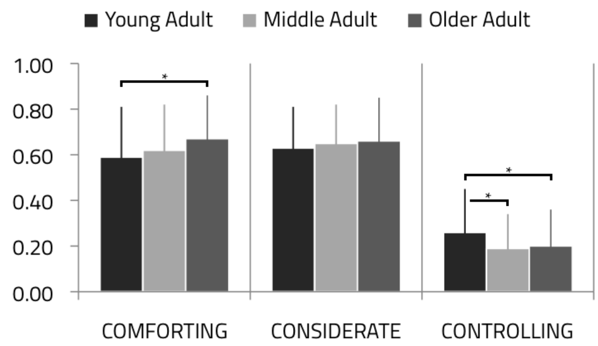


Figure 3: Main effects of participant *age*. Asterisks indicate significant contrasts.

3.2.3 Interactions

Furthermore at odds with Experiment I (where the $gender \times politeness$ eclipsed many of the main effects of politeness), there were no significant or even marginally significant interaction effects here. Specifically, in the context of the more naturalistic interactions, the use of polite speech seemed to be effective for both female *and* male participants. This suggests that, while female participants appear to be particularly sensitive to verbal communication (as evidenced by their ratings across both the more naturalistic Experiment II *and* Experiment I), male participants may be more sensitive to *consistency* in verbal and nonverbal communicatory cues.

3.3 Discussion

3.3.1 Summary of Present Findings & Implications

In this follow-up investigation, we explored whether our previous findings in Experiment I – that a robot's use of polite speech improves perceptions (and, that women respond more positively towards such robots) – hold given more naturalistic interaction scenarios (i.e., human-robot interactions in which the robot is animated).

Here we observed that the results, for the most part, reflect those of the previous experiment (despite E1's lack of movement in the shown video interactions). Specifically, the politeness manipulation again resulted in lower ratings of the robot as *controlling* and higher ratings of the robot as being *considerate* and *comforting* (see Figure 4, left). This lends further support of politeness as an effective tool for facilitating more positive responding towards robots (at least for natural language interactions in advice-giving scenarios).

Similarly, participants' *gender* again exerted a main effect on *all* dependent measures: how *comforting*, *considerate*, and *controlling* the robot was perceived as being (see Figure 4, right), as well as how *difficult* both the task and interaction seemed and participants' *interest* in interacting with the depicted robot. In particular, women rated the robots more positively than did male participants as was observed in Experiment I. While this remains in contradition with prior work showing that men respond more positively towards robots than women (e.g., [15, 18, 20]), it nevertheless lends further support towards the methodological implication that gender is a relevant consideration for HRI studies.

Moreover, the results of the present investigation indicate that observatory perspectives of human-robot interactions are not substantially influenced by the robot's animacy. This suggests that simplistic

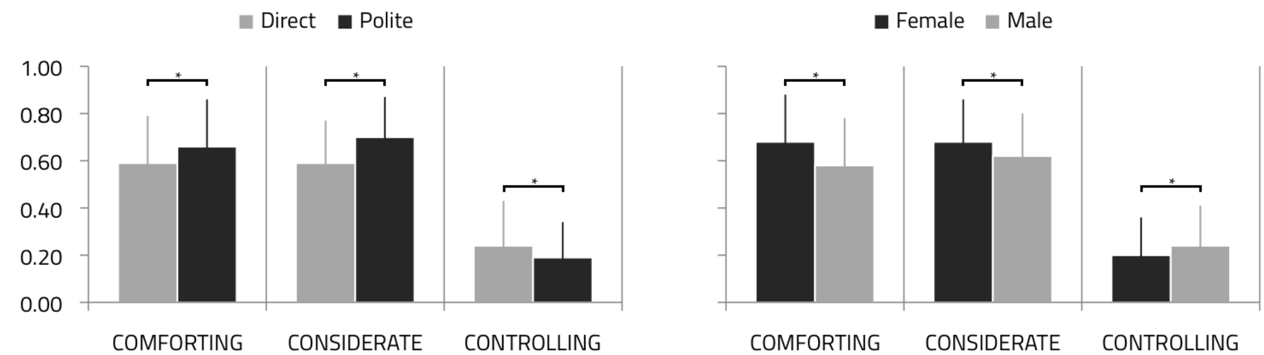


Figure 4: Main effects of *politeness* (left) and participant *gender* (right) on perceptions of the robot as *comforting*, *considerate*, and *controlling*. Dark bars emphasize factors yielding more positive outcomes (polite speech, female participants). All contrasts are significant.

depictions of human-robot interactions, such as in Experiment I, may suffice to investigate perceptions of certain robot behaviors (e.g., a robot's politeness, as perceived by female observers).

However, key differences in findings between the two experiments also underscore the necessity of considering perceptions in more realistic interaction scenarios. Specifically, unlike in Experiment I, Experiment II showed no interactions between any of the three IVs. For example, in the context of the more naturalistic interactions, the use of polite speech was effective at improving ratings regardless of the participant's gender. Whereas, in Experiment I, polite speech was only effective at improving *female participants'* ratings (while male participants of Experiment I were not receptive – the use of politeness modifiers, in the absence of the idling and attention sharing movements, did not improve ratings). This suggests that, while women appear to be particularly sensitive to verbal communication (as evidenced by their ratings across both Experiment I and the more naturalistic Experiment II), men may be more sensitive to *consistency* in verbal and nonverbal communicatory cues. Thus, the findings may imply a need for coherence between a robot's verbal and nonverbal communication (e.g., [10]).

In addition, the present experiment showed a slight influence of age on perceptions of *comfort* with the robot and how *controlling* it seemed (see Figure 3), whereas E1 showed no significant effects owing to participants' age. These effects are somewhat difficult to interpret, however, as it is unclear what aspects of the more realistic interaction would cause the standard university-aged participants (relative to the older adults) to here indicate less comfort with the robot and rate it as more controlling.

3.3.2 Limitations & Future Directions

Here we undertook further investigation of perceptions of robot politeness and potential modulatory factors. Our approach tested a few simple behaviors to assess the influence (or lack thereof) of a robot's movement. In particular, the presentation of human-robot interactions that were more naturalistic (i.e., mimic attention-sharing and idling behaviors exhibited in equivalent human-human interactions) compliments our previous study, which lacked the same degree of social realism. This contribution here is significant because it replicates the influences of politeness of both prior work and our own Experiment I. Further, it sheds light on how subject-based factors (i.e., age and gender) can yield more positive social evaluations. However, as with the previous study, our approach still has its limitations.

In particular, we explored here only a small subset of human-inspired movements. Thus, it is not possible to conclusively say that movement (of any kind) is effective for improving interactions or perceptions thereof. There are substantially more possibilities to try, such as gaze aversion (e.g., [1]) or gesturing (e.g., [11]) to name a few. To determine what extent certain types nonverbal communicatory mechanisms influence perceptions, future work might consider independently manipulating several types of movements, rather than the movement/no-movement meta comparison we made here.

4 GENERAL DISCUSSION

4.1 General Findings & Implications

As expected, Experiment I confirms prior indications that, at least in 3rd-person observation of pre-recorded human-robot interactions ([21, 22]), a robot's use of politeness modifiers in its speech is perceived more favorably relative to a robot that uses more direct speech (e.g., [14, 19, 21, 22]). This is reflected by participants ratings of the polite robot instructors as more *comforting* and *considerate*, and less *controlling* than the robots that were more direct. Moreover, the implications of politeness hold, even for a population that is highly diverse in terms of the socio-demographic factors of education, geographical location, age, and gender. Furthermore, we observed additional validation of the effects owing to a robot's politeness in Experiment II. Thus, consistent with prior indications ([21, 22]), the persistence of effects due to politeness – given the broader population sampling – demonstrate the benefit to using politeness modifiers when a robot communicates with natural language.

The results observed across the two studies further underscore an important methodological consideration – namely, gender – for evaluation of human-robot interactions. Specifically, we found a gender-based divide in the efficacy of the politeness manipulation in both experiments showing that a robot's use of politeness modifiers in its speech is most (and in Experiment I, only) effective for female participants. That is, here women rated polite robots significantly better than those that are more direct, and moreover, their ratings of polite robots are significantly higher than men's ratings of the same robots. Furthermore, the two studies suggest that men are sensitive to consistency in communicatory cues, and more importantly, they are not receptive to polite speech alone. These findings demonstrate the importance of considering gender – either as a systematic manipulation or as a covariate – in the analysis of human-robot interactions.

4.2 General Limitations & Future Work

Our approach to understanding perceptions of polite robots contributes a simple online task to assess the modulatory influences of various situational factors. We emphasize the benefit that the online forum serves for obtaining data with broad socio-demographics versus in-laboratory studies which are limited to smaller and more homogeneous participant populations. This lends the ability towards replicating previously indicated influences of politeness and understanding how such findings might transfer to the general population.

However, we wish to also underscore the limitations of this type of assessment. Despite the benefits to online studies, the results cannot be immediately applied to actual human-robot interactions involving co-located, direct participation, as the present work was conducted from a remote and observatory position (relative to the depicted interactions). Hence, whether (and if so, the extent to which) these findings generalize and apply to in-person, direct interactions with a co-located embodied agent motivates further investigation.

Further, we stress that these findings are preliminary and of limited weight. In particular, we note the small effect sizes observed across both studies. Between the two experiments, the effect sizes reached at most a medium qualification with the influence of politeness on perceptions of the robot as *considerate* ($\eta^2 = .11$ in the more naturalistic interaction scenario of Experiment II, and $\eta^2 = .06$ in Experiment I). Gender also showed an effect of close to a medium size on ratings of *comfort* ($\eta^2 = .06$). However, the size of other effects observed (e.g., due to age) is small ($\eta^2 \leq .03$). Thus, relative to other factors (e.g., the robot's appearance), the robot's politeness and the person's age/gender may be of little importance. While the present work yields implications for both the design of robotic agents and how to evaluate them, future work might consider how relevant gender and politeness are in other contexts or in contrast to other factors.

5 CONCLUSIONS

The primary aim of this research was to investigate whether previous results about human observers' preferences for polite robot speech over more direct speech in an robot instructor would hold for a wider participant demographic, which we were able to confirm. A secondary aim was to explore the modulatory influences of a person's age and gender on perceptions of the robot. Here we obtained several new and important gender effects that hint at a complex interplay of the interaction observers' gender with the observed robot's behavior, which warrants further investigation to elucidate the causal mechanisms responsible for the gender-based differences. Further, owing to a limitation of the design of our first experiment, we explored peoples' perceptions given a more realistic interaction scenario which additionally confirmed the influence of both politeness and gender. These findings are particularly important for the design of future autonomous agents, robotic or virtual, because their success could significantly depend on their ability to adapt, such as to gender-specific expectations of their interactants.

REFERENCES

[1] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu, 'Conversational gaze aversion for humanlike robots', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 25–32, (2014).

[2] Christoph Bartneck, Tomohiro Suzuki, Takayuki Kanda, and Tatsuya Nomura, 'The influence of peoples culture and prior experiences with aibo on their attitude towards robots', *AI & Society*, **21**(1), 217–230, (2007).

[3] Charles R Crowell, Michael Villano, Matthias Scheutz, and P. Schermerhorn, 'Gendered voice and robot entities: perceptions and reactions of male and female subjects', in *Proceedings of the IEEE International Conference on Intelligent Robots and Systems*, pp. 3735–3741, (2009).

[4] Juan Fasola and Maja J. Mataric, 'A socially assistive robot exercise coach for the elderly', *Journal of Human-Robot Interaction*, **2**(2), 3–32, (2013).

[5] Priska Flandorfer, 'Population ageing and socially assistive robots for elderly persons: the importance of sociodemographic factors for user acceptance', *International Journal of Population Research*, (2012).

[6] ME Foster, A Gaschler, M Guiliani, A Isard, M Pateraki, and RPA Petrick, 'Two people walk into a bar: Dynamic multi-party social interaction with a robot agent', in *International Conference on Multimodal Interaction (ICMI)*, pp. 3–10, (2012).

[7] Bruce Fraser, 'Conversational mitigation', *Journal of Pragmatics*, **4**, 341–350, (1980).

[8] Kerstin Sophie Haring, David Silvera-Tawil, Yoshio Matsumoto, Mari Velonaki, and Katsumi Watanabe, 'Perception of an android robot in japan and australia: A cross-cultural comparison', in *Social Robotics*, 166–175, (2014).

[9] Lewis Hassell and Margaret Christensen, 'Indirect speech acts and their use in three channels of communication', in *Communication Modeling*, p. 9, (1996).

[10] J Hodgins, S Jörg, C O'Sullivan, SI Park, and M Mahler, 'The saliency of anomalies in animated human characters', *ACM Transactions on Applied Perception (TAP)*, **7**(4), 1–14, (2010).

[11] Heeyoung Kim, Sonya S Kwak, and Myungsuk Kim, 'Personality design of sociable robots by control of gesture design factors', in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pp. 494–499, (2008).

[12] VA Kulyukin, 'On natural language dialogue with assistive robots', in *HRI*, pp. 164–171, (2006).

[13] I Han Kuo, Joel Marcus Rabindran, Elizabeth Broadbent, Yong In Lee, Ngaire Kerse, RMQ Stafford, and Bruce A MacDonald, 'Age and gender factors in user acceptance of healthcare robots', in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication*, pp. 214–219, (2009).

[14] Min Kyung Lee, Sara Kiesler, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski, 'Gracefully mitigating breakdowns in robotic services', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 203–210, (2010).

[15] Bilge Mutlu, Steven Osman, Jodi Forlizzi, Jessica Hodgins, and Sara Kiesler, 'Task structure and user attributes as elements of human-robot interaction design', in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication*, pp. 74–79, (2006).

[16] Tatsuya Nomura and Kazuma Saeki, 'Effects of polite behaviors expressed by robots: A psychological experiment in japan', *International Journal of Synthetic Emotions*, **1**(2), 38–52, (2010).

[17] Tatsuya Nomura and Satoru Takagi, 'Exploring effects of educational backgrounds and gender in human-robot interaction', in *Proceedings of the International Conference on User Science and Engineering*, pp. 24–29, (2011).

[18] Natalia Reich and Friederike Eyssel, 'Attitudes towards service robots in domestic environments: The role of personality characteristics, individual interests, and demographic variables', *Journal of Behavioral Robotics*, **4**(2), 123–130, (2013).

[19] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Marhaba, how may I help you?: Effects of politeness and culture on robot acceptance and anthropomorphization', in *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction*, pp. 74–81, (2014).

[20] Paul Schermerhorn, Matthias Scheutz, and Charles R. Crowell, 'Robot social presence and gender: Do females view robots differently than males?', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 263–270, (2008).

[21] Megan Strait, Cody Canning, and Matthias Scheutz, 'Let me tell you! Investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality, and distance', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 479–486, (2014).

[22] Cristen Torry, Susan R. Fussell, and Sara Kiesler, 'How a robot should give advice', in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp. 275–282, (2013).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Perception of Artificial Agents and Utterance Friendliness in Dialogue

Sascha Griffiths[1] and **Friederike Eyssel**[2] and **Anja Philippsen** and **Christian Pietsch** and **Sven Wachsmuth**[3]

Abstract. The present contribution investigates the construction of dialogue structure for the use in human-machine interaction especially for robotic systems and embodied conversational agents. We are going to present a methodology and findings of a pilot study for the design of task-specific dialogues. Specifically, we investigated effects of dialogue complexity on two levels: First, we examined the perception of the embodied conversational agent, and second, we studied participants' performance following HRI. To do so, we manipulated the agent's friendliness during a brief conversation with the user in a receptionist scenario.

The paper presents an overview of the dialogue system, the process of dialogue construction, and initial evidence from an evaluation study with naïve users (N = 40). These users interacted with the system in a task-based dialogue in which they had to ask for the way in a building unknown to them. Afterwards participants filled in a questionnaire. Our findings show that the users prefer the friendly version of the dialogue which scored higher values both in terms of data collected via a questionnaire and in terms of observations in video data collected during the run of the study.

Implications of the present research for follow-up studies are discussed, specifically focusing on the effects that dialogue features have on agent perception and on the user's evaluation and performance.

1 Introduction

Research within the area of "language and emotion" has been identified as one key domain of innovation for the coming years [40, 20]. However, with regard to human-machine communication, we still need better speech interfaces to facilitate human-robot interaction (HRI) [30, 31]. Previous work on human-human communication has already demonstrated that even small nuances in speech have a strong impact on the perception of an interlocutor [1, 38].

In the present work, we have therefore focused on the role of dialogue features (i.e., agent verbosity) and investigated their effects on the evaluation of an embodied conversational agent (ECA) and the user performance. We designed a receptionist scenario involving a newly developed demonstrator platform (see Section 3.2) that offers great potential for natural and smooth human-agent dialogue. To explore how to model dialogues efficiently within actual human-robot interaction we relied on a Wizard-of-Oz paradigm [16, 17].

[1] Queen Mary University of London, UK, email: sascha.griffiths@qmul.ac.uk
[2] New York University, Abu Dhabi, email: fae5@nyu.edu
[3] Bielefeld University, Germany, email: anja.philippsen, christian.pietsch, sven.wachsmuth@uni-bielefeld.de

This HRI scenario involved an embodied conversational agent which served as a receptionist in the lobby of a research center. A similar set-up has been realized in previous studies [2, 24, 25]. Moreover, we draw from existing research on dialogue system design [33] and the acceptance of artificial agents [13, 22].

The question that we seek to answer arises frequently during the implementation of a robot scenario (such as this receptionist scenario) [26], and can also be phrased as how the system should verbalize the information that it is supposed to convey to the user. Obviously, a script has to be provided that covers the necessary dialogue content. The relevant issue is that each utterance can be phrased in a number of ways. This brings up several follow-up questions such as: *Can the perceived friendliness of an agent be successfully manipulated? Is the proposed script a natural way of expressing the intended meaning? Are longer or shorter utterances favourable? How will the user respond to a given wording? Will the script elicit the appropriate responses from the user?*

For the purpose of investigating these questions, we will first discuss related literature and relevant theoretical points. The following section will describe the system. We then turn to the dialogue design and first empirical evidence from a user study.

2 Dialogue Complexity and Perception of Artificial Agents

Obviously, the issue of how to realize efficient dialogue in HRI has been of interest to many researchers in the area of human-machine interaction and principles of natural language generation are generally well understood [39]. However, this is less so the case when taking into account communication patterns between humans and embodied conversational agents and robots.

2.1 Dialogue Complexity and Social Meaning

As Richard Hudson notes, "social meaning is spread right through the language system" [23]. Thus, there is a clear difference between interactions if one commences with the colloquial greeting "*Hi!*" versus one initiated with a more polite "*Good Morning*". However, this does not only concern peripheral elements of language such as greetings, but also syntax. Hudson uses the following example to illustrate this:

1. *Don't you come home late!*
2. *Don't come home late!*

Both sentences differ in terms of syntax and their social meaning. The syntax varies as the first sentence explicitly refers to the subject, whereas the second sentence does not. The first sentence in the example also appears more threatening in tone than the latter. These subtle differences in the statements' wording lead to a fundamentally different interpretation. Analogously, we assume that in human-agent dialogue subtle manipulations of aspects of that dialogue can result in changes in agent perception. Concretely, we will investigate the role of this kind of linguistic complexity [11] within human-machine interaction.

The impact of changing a dialogue with respect to the social meaning communicated has already been tested in the REA (an acronym for "Real Estate Agent") system [9, 5]. In a study [4] of users' perception of different versions of REA's behaviour, a "normal REA" was tested against an "impolite REA" and a "chatty REA". Results indicated that in the condition in which REA was able to produce a small amount of small talk REA was judged more likeable by participants. In further studies with the system the authors concluded that the interpersonal dimension of interaction with artificial agents is important [8]. It has been shown that implementing a system which achieves task goals and interpersonal goals as well as displaying its domain knowledge can increase the trust a user will have in a system [3]. Cassell [7] also argues that equipping artificial agents with means of expressing social meaning not only improves the users' trust in the domain knowledge that such systems display but also improves interaction with such systems as the users can exploit more of their experience from human-human dialogue.

2.2 Interaction Patterns

The dialogue flow used in the present study was implemented with PaMini, a pattern-based dialogue system which was specifically designed for HRI purposes [32] and has been successfully applied in various human-robot interaction scenarios [35, 36, 37]. The dialogue model underlying the present system (see Section 3.1) is therefore based on generic interaction patterns [33]. Linguistically speaking these are adjacency pairs [29, 10]. In these terms, a dialogue will consist of several invariant elements which are sequentially presented as pairs with one interlocutor uttering one half of the pair in his turn and the other interaction partner responding with an appropriate response.

The full list of generic interaction patterns which are distinguished according to their function given by Peltason et al. [34] includes the following utterance categories: *Greeting, Introducing, Exchanging pleasantries, Task transition, Attracting attention, Object demonstration, Object query, Listing learned objects, Checking, Praising, Restart, Transitional phrases, Closing task, Parting*.

For all these dialogue tasks one can see the interaction as pairs of turns between interlocutors. Each partner has a certain response which fits to the other interlocutor's utterance. Examples of this kind of interaction can be found in Table 1.

Table 1. Examples of adjacency pairs in human-robot interaction (adapted from [34])

Purpose	Example interaction
Greeting	*User: Hello, Vince.* *Robot: Hi, hello.*
Introducing	*User: My name is Dave.* *Robot: Hello, Dave. Nice to meet you.*
Object query	*Robot: What is that?* *User: This is an apple.*
Praising	*User: Well done, Vince.* *Robot: Thank you.*

The problem one faces is that while such dialogues are based on generic speech acts, there is the remaining problem of how the individual items need to be worded. Winograd [46] distinguishes between the ideational function and interpersonal function of language. The ideational function can loosely be understood as the propositional content of an utterance whereas the interpersonal function has more to do with the context of an utterance and its purpose.

3 System Architecture

In the following, we present the system which was constructed both as a demonstrator and as a research platform. We will present the entire set-up which includes an ECA, Vince [42], and a mobile robot platform, Biron [21]. Both of these use the same dialogue manager but only the ECA has been used in this pilot study.

Figure 1 illustrates the architecture of the complete system in autonomous mode. Communication between the components is mainly implemented using the XML-based XCF framework and the Active Memory structure [47]. Three memories are provided for different kinds of information: The short term memory contains speech related information which is inserted and retrieved by the speech recognizer, the semantic processing unit and the dialogue manager. The visual memory is filled by the visual perception components, it contains information about where persons are currently detected in the scene.

The system is designed to provide the visitor verbally with information, but also to guide them to the requested room if necessary[4]. For this purpose, the agent Vince communicates information about the current visitor and his or her needs to the mobile robot Biron via a shared (common ground) memory.

Although Biron is omitted in the present study to reduce complexity, we present the complete system, as Vince and Biron use the same underlying dialogue system. Note that the study could have been conducted also with Biron instead of Vince. Such a study is subject to future work.

3.1 Dialogue Manager

The dialogue manager plays a central role in the overall system as it receives the pre-processed input from the user and decides for adequate responses of the system. A dialogue act may also be triggered by the appearance of persons in the scene as reported by the visual perception component.

Speech input from the user is recognized using the ISR speech recognizer based on ESMERALDA [14]. The semantic meaning is extracted via a parsing component which is possible due to the well defined scenario. Additionally, this component retrieves missing information from an LDAP server that the human might be interested in (e.g. office numbers). The dialogue manager PaMini [35, 36, 37] is based on finite state machines which realize interaction patterns for different dialogue situations as described in Section 2.2. Patterns are triggered by the user or by the robot itself (mixed-initiative). The dialogue component sends the selected response and possibly gesture instructions to the Vince system which synchronizes the speech output and the gesture control internally [28, 27]. Exploiting the information from the visual perception component, Vince attends to the current visitor via gaze following [24].

Biron incorporates a separate dialogue which is coupled with the Vince dialogue. The Biron dialogue at the moment receives input

[4] A short video demonstration of the scenario is provided in this CITEC video: `http://www.youtube.com/watch?v=GOz_MsLel1Y#t=4m32s`. Accessed: March 2, 2015

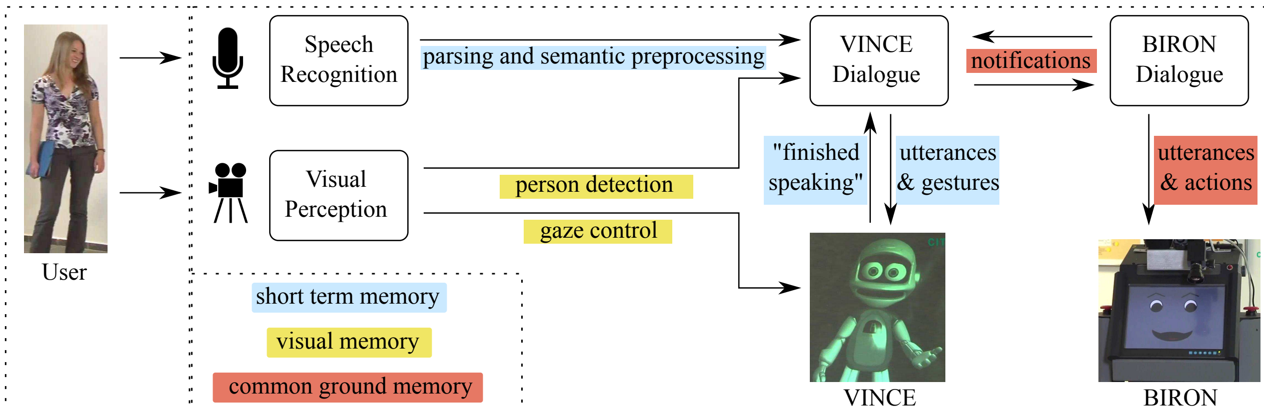


Figure 1. Overview of the architecture of the system in autonomous mode. The colors of the three memories indicate which information is stored in which memory. See Section 3.1 for a thorough description of the information flow.

solely from the Vince dialogue component (not from the user) and communicates the current state to the user. If the visitor wishes, Vince calls Biron and orders him to guide the visitor to the requested room. This feature is currently limited to offices on the ground floor, if visitors are looking for a room on the first or second floor, Biron guides them to the elevator and provides them with information about how to find the room on their own.

3.2 Demonstrator Platform

The embodied conversational agent Vince is installed on a workstation. An Apple Mac Mini is used for this purpose. The system runs a UNIX based operating system (Linux Ubuntu 10.04 32bit). The user interface is controlled by a wireless bluetooth mouse and keyboard or via remote access. The ECA is displayed on a holographic projection screen (i.e. a HoloPro Terminal[5]) in order to achieve a high degree of perceived embodiment. A microphone records speech input and video data are recorded using two cameras. Two loudspeakers are connected to the Mac Mini workstation to provide audio output.

4 Study Design and Realisation

We set up a simplified version of the CITEC Dialogue Demonstrator for the purpose of the study. One difference is that we do not make use of the mobile robot Biron here. Secondly, we rely on Wizard-of Oz teleoperation [12, 45] to trigger interaction patterns by means of a graphical user interface that was designed for our case study.

4.1 Preparation of Dialogues

The dialogues were prepared bottom-up. We tried to leave as little as possible to design by the researchers or a single researcher.

To investigate human-machine dialogue in the context of a receptionist scenario, we initially simulated such dialogues between two human target persons who were given cards which described a particular situation (e.g. that a person would be inquiring about another persons office location).

We recorded two versions of eight dialogues with the two participants, who were asked to take the perspective of a receptionist or a visitor, respectively. The dialogues were then transliterated by a third party who had not been involved in the staged dialogues.

To model the receptionist turns, we extracted all phrases which were classified as greetings, introductions, descriptions of the way to certain places and farewells. We then constructed a paper-and-pencil pre-test in order to identify a set of dialogues that differed in friendliness. 20 participants from a convenience sample were asked to rate the dialogues with regard to perceived friendliness using a 7-point Likert scale.

These ratings were used as a basis to construct eight sample dialogues which differed both in friendliness and verbosity. In a subsequent online pre-test, the sample dialogues were embedded in a cover-story that resembled the set-up of our WoZ scenario.

We used an online questionnaire to test how people perceived these dialogues. On the start screen participants were presented with a picture of the embodied conversational agent Vince and told that he would serve as a receptionist for the CITEC building. On the following screens textual versions of the eight human-agent dialogues were presented. Participants were asked to rate these dialogues with regard to friendliness in order to identify dialogues that would be perceived as either low or high in degree of perceived friendliness of the interaction.

The dialogue with the highest rating for friendliness and the dialogue with the lowest rating for friendliness were then de-composed into their respective parts and used in the main study. The two dialogue versions are presented in Table 2.

4.2 Study

In the main study, the participants directly interacted with the ECA which was displayed on a screen (see Figure 1).

We recruited students and staff at the campus of Bielefeld University to participate in our study on “human-computer interaction”. 20 male and 20 female participants ranging in age from 19 to 29 years (M = 23.8 years, SD = 2.36) took part in the study. Before beginning their run of the study, each participant provided informed consent. Each participant was then randomly assigned to one of two conditions in which we manipulated dialogue friendliness.

The study involved two research assistants (unbeknownst to the participants). Research assistant 1 took over the role of the “wizard” and controlled the ECA’s utterances, while research assistant 2 interacted directly with the participants.

[5] `http://www.holopro.com/de/produkte/holoterminal.html` Accessed: March 2, 2015

Table 2. Friendly and neutral dialogue version

Dialogue Act	Neutral version	Friendly version
Greeting	Hallo *Hello*	Guten Tag, kann ich Ihnen helfen? *Good afternoon, how can I help you?*
Directions	Der Fragebogen befindet sich in Q2-102. *The questionnaire is located in Q2-102.*	Der Fragebogen befindet sich in Raum Q2 102. Das ist im zweiten Stock. Wenn Sie jetzt zu Ihrer Rechten den Gang hier runter gehen. Am Ende des Gangs befinden sich die Treppen, diese gehen Sie einfach ganz hoch und gehen dann durch die Feuerschutztür und dann ist der Raum einfach geradeaus. *The questionnaire is located in room Q2-102. That is on the second floor. If you turn to your right and walk down the hallway. At the end of the floor you will find the stairs. Just walk up the stairs to the top floor and go through the fire door. The room is then straight ahead.*
Farewell	Wiedersehen. *Goodbye.*	Gerne. *You are welcome.*

Following the Wizard-of-Oz paradigm, research assistant 1 was hidden in the control room and controlled the ECA's verbalisations using a graphical user interface. A video and audio stream was transmitted from the dialogue system to the control room. The "wizard" had been trained prior to conducting the study to press buttons corresponding to the "Dialogue Acts" as shown in Table 2. Importantly, research assistant 1 only knew the overall script (containing a greeting, a description of the route to a room and a farewell), but was blind to the authors' research questions and assumptions.

To initiate the study, research assistant 1 executed "Greeting A" or "Greeting B", depending on whether the "friendly" or "neutral" condition was to be presented, then proceeded to pressing "Directions A" or "Directions B" and finally "Farewell A" and "Farewell B" once the user had reacted to each utterance.

The users then had to follow the instruction given by the agent. Research assistant 2 awaited them at the destination where they had to fill in a questionnaire asking for their impressions of the interaction.

The questionnaire investigated whether differential degrees of dialogue complexity would alter the perception of the artificial agent with respect to a) warmth and competence [15], b) mind attribution [19], and c) usability (system usability scale *SUS*) [6]. We consider these question blocks as standard measures in social psychology and usability studies.

The questionnaire was comprised of three blocks of questions. These do to some extent correspond to the four paradigms of artificial intelligence research listed in Russell & Norvig [41]: "thinking humanly", "acting humanly", "thinking rationally" and "acting rationally". As we were only looking at perception of the artificial agent, we did not look into "thinking rationally". However, warmth and competence are used in research on anthropomorphism, which one can regard as a form of "acting humanly". Mind perception can be related to "thinking humanly". Usability (SUS) is a form of operationalising whether an artificial agent is acting goal driven and useful which holds information on whether it is "acting rationally".

The first block of the questionnaire included four critical items on warmth, and three critical items on competence, as well as nine filler items. The critical questions asked for attributes related to either warmth, such as "good-natured", or competence, such as "skillful".

The second block consisted of 22 questions related to mind perception. These questions asked the participants to rate whether they believed that Vince can be attributed mental states. A typical item is the question whether Vince was capable of remembering events or whether he is able to feel pain.

Finally, the SUS questionnaire consisted of 10 items directly related to usability. Participants were asked question such as whether they found the system easy to use.

Upon completion of the questionnaire, participants were debriefed, reimbursed and dismissed.

5 Results

In the following, two types of results are reported. In Section 5.1, we present results from the questionnaire, in Section 5.2, we present initial results from video data recorded during the study.

5.1 Questionnaire Responses

As aforementioned, 7-point Likert scales (for the warmth, competence and mind question blocks) and a 5-point Likert scale for the SUS questions block) were used to measure participants responses to the dependent measures. For each dependent variable, mean scores were computed with higher values reflecting greater endorsement of the focal construct. Values for the four blocks of questions were averaged for further analysis. The results for the questionnaire are shown in Figure 2.

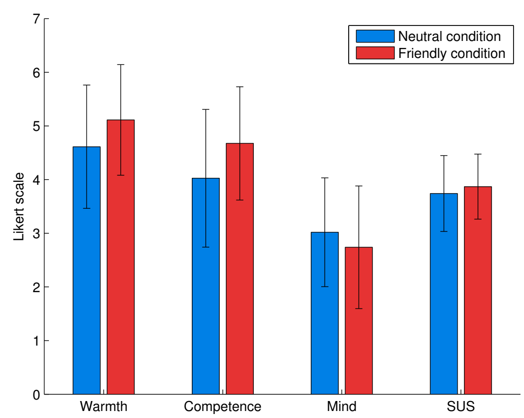


Figure 2. Mean response values for the questionnaire question sets. The mean for the dependent variables warmth, competence, mind and SUS are compared for the two categories neutral (blue) and friendly (red).

5.1.1 *Warmth*

The mean values for the warmth question set can be seen in Figure 2. It can be notices that the values for the friendly condition are mostly higher than for the neutral condition. The descriptive statistics confirm this. The friendly condition has a maximum value of 7 and a minimum value of 3.25 whereas the neutral condition has a maximum value of 6.75 and a minimum value of 2.25. The mean of the friendly condition is M = 5.11 (SD = 1.14) and the mean of the neutral condition is M = 4.61 (SD = 1.14). The mean values suggest that within the population on which our system was tested the friendly condition is perceived warmer than the neutral condition.

5.1.2 Competence

Similarly, the values for the friendly condition are mostly higher than for the neutral condition. The descriptive statistics confirm this. The friendly condition has a maximum value of 7 and a minimum value of 2.75 whereas the neutral condition has a maximum value of 6.25 and a minimum value of 1.5. The mean of the friendly condition is M = 4.68 (SD = 1.05) and the mean of the neutral condition is M = 4.02 (SD = 1.28). The standard deviation shows that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested the friendly condition is perceived more competent than the neutral condition.

5.1.3 Mind Perception

As Figure 2 shows, the ECA is perceived slightly higher on mind perception in the neutral condition than in the the friendly condition. The neutral condition has a maximum value of 4.9 and a minimum value of 1.32 whereas the friendly condition has a maximum value of 4.93 and a minimum value of 1.09. However, the mean of the neutral condition is M = 3.02 (SD = 1.01) whereas the mean of the friendly condition is M = 2.74 (SD = 1.14). The standard deviation suggests that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested in the friendly condition the participants attributed less mind to the ECA than the neutral condition.

5.1.4 System Usability Scale (SUS)

The values on the system usability scale are slightly higher in the friendly condition than in the neutral condition. The friendly condition has a maximum value of 4.7 and a minimum value of 2.7 whereas the neutral condition has a maximum value of 4.9 and a minimum value of 2.5. The mean of the friendly condition is M = 3.87 (SD = 0.61) and the mean of the neutral condition is M = 3.74 (SD = 0.71). The standard deviation suggests that there is more variation in the values for the neutral condition. The mean values overall suggest that within the population on which our system was tested the friendly condition was rated slightly more usable than the neutral condition.

5.2 Further Observations

Further observations that could be made on the dialogue level resulted from the analysis of the video data collected during the runs of the study. The dialogues were transcribed and inspected by one student assistant[6] trained in conversation analysis [18]. The purpose of this was to examine the dialogues to find out whether there were any particular delays in the dialogues and whether participants conformed to the script or not.

[6] Taking this line of research further, we would use two annotators and check for agreement between them. However, this was beyond the scope of the current contribution.

5.2.1 Alignment

We looked at the mean utterance length (MUL) of the participants in interaction with the ECA. We take this as an indicator of how participants align their verbalisations with the agent's verbalisations. The differences between the two conditions can be seen in in Figure 3, the values for the friendly condition are mostly higher than for the neutral condition.

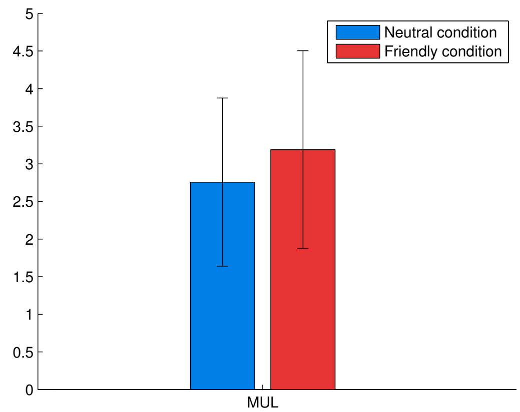


Figure 3. The mean utterance length averaged over the two conditions. The friendly condition has a slightly higher mean value than the neutral condition.

The descriptive statistics confirm this. The friendly condition has a maximum value of 5.5 and a minimum value of 1 whereas the neutral condition has a maximum value of 5.25 and a minimum value of 1. The mean of the friendly condition is M = 3.12 (SD = 1.31) and the mean of the neutral condition is M = 2.76 (SD = 1.11). The standard deviation suggests that there is more variation in the values for the friendly condition. The mean values overall suggest that within the population on which our system was tested the friendly condition showed more alignment with the ECA's MUL than the neutral condition.

5.2.2 Irregularities

The video data were reviewed and four types of noticeable effects on the dialogue were determined:

1. Participants returning because they did not understand or forget the ECA's instructions (22.5%, see Section 5.2.3),
2. deviations from the script, i.e. participants trying to do small talk with the ECA (5%, see Section 5.2.4),
3. timing difficulties causing delays in the interaction (25%), and
4. other ways in which the script was altered in small ways (22.5%, e.g. mismatches between the ECA's utterances and the participants utterances).

The overall number of irregularities accumulated across the two categories is summarized in Table 3. In interactions with the neutral condition irregularities can be observed in 75% of the cases, while in the friendly condition only 50% of the interactions show irregularities.

Table 3. Overview of occurred irregularities in the neutral and friendly condition.

	Neutral	Friendly
No irregularities	5	10
Irregularities occur	15	10

5.2.3 *Clarity of instructions*

Out of the 40 interactions in 9 cases (22.5%) the participants returned because they realized that they could not remember the room number correctly. Out of these the majority, namely 6, were in the neutral condition. Three participants came back for a second short interaction with Vince in the friendly condition.

5.2.4 *Small talk*

Only two participants (5%) deviated from the script of the dialogue by attempting to do small talk with Vince. Both of these were in the friendly condition. One participant asked the ECA for its name. Another participants tried three deviating questions on Vince during the interaction. The first question was "How are you?", the second "What can you tell me?", and finally the ECA was asked whether they were supposed to actually go to the room after the instructions were given.

6 Discussion

In reporting our results we concentrated on the descriptive statistics and no attempt will be made to generalize beyond this population. Within this first pilot study with the current demonstrator we tried to assess whether manipulating the degree of perceived friendliness has an effect on the interaction.

We now return to the questions asked in the introduction, the main question being how the manipulation affected the interaction between the user and the artificial agent.

6.1 Can the perceived friendliness of an agent be successfully manipulated?

We obtained slightly higher values regarding the perceived warmth in the friendly condition as opposed to the neutral condition. The differences are very small, though. The descriptive statistics point towards a "friendly" version of the dialogue actually being perceived as more friendly by the user. We propose that this will make users more willing to use the services the system can provide. Thus, further research into "friendly agents" seems a productive agenda.

The friendliness level also suggested higher ratings for competence, despite the fact that the friendly dialogue actually led to more misunderstandings. This failure was not reflected in the users judgements directly. Also, participants seem to prefer interacting with the friendly agent.

6.2 Is the proposed script a natural way of expressing the intended meaning?

The results which the video data analysis presented indicate that actually the majority of interactions conducted within this study were smooth and there were no noticeable deviations from the overall "script" in most dialogues. The operator was able to conduct most of the dialogues with the use of just a few buttons. This suggests that one can actually script dialogues of this simple nature quite easily.

However, the wording is crucial and the results suggest that the friendly version of the dialogue is more amicable to clarity. Only three participants did not fully understand or remember the instructions whereas twice as many had to ask for the room a second time in the neutral condition.

6.3 Are longer or shorter utterances favourable?

In a task-based dialogue the artificial agent will ideally demonstrate its knowledge and skill in a domain. However, the pilot-study did not find a very high difference between the two conditions regarding the competence question. The descriptive statistics, however, suggest that the longer utterances in the friendly dialogue received higher competence ratings.

Converse to the prediction, mind perception was slightly higher for the neutral dialogue, though. Thus, the friendly agent is not necessarily perceived as more intelligent by the user.

However, the longer utterances in the friendly version of the dialogue received higher ratings with respect to usability. Also, fewer participants had to come back and ask for the way again in a second interaction in the friendly condition. This suggests that the longer version of the dialogue better conveyed the dialogue content than the neutral version.

6.4 How does the user respond to a given wording?

In the friendly condition, users used longer utterances themselves when speaking to the friendly version of the ECA with more verbose verbalisations. This shows that the participants do align their speech with that of the artificial agent.

One can also tell from the video analysis that only in the friendly condition participants were motivated to further explore the possibilities the system offers. Two participants decided to ask questions which went beyond the script.

6.5 Will the script elicit the appropriate responses from the user?

Participants found it easy to conform to the proposed script. There was only a low percentage of participants who substantially deviated from the script and stimuli presented by the ECA (5% tried to do small talk with the agent). Most dialogues proceeded without the participants reacting in unanticipated ways and only a small percentage of participants failed to extract the relevant information from the verbalisations of the artificial agent.

7 Conclusion

We presented a pilot-study in which participants were confronted with dialogue exhibiting different degrees of friendliness.

While maintaining the same ideational function (see Section 2.2 above) we changed the interpersonal function of the dialogue by using sentences which were obtained through a role-playing pre-study and then rated by participants according to their friendliness.

The obtained dialogues (a friendly and a neutral version) were presented to participants in interaction with an ECA which was implemented via generic interaction patterns. Participants filled in a questionnaire after the interaction which was analysed along with further observational data collected during the study.

The results point towards higher perceived warmth, higher perceived competence and a greater usability judgement for the ECA's performance in the friendly condition. However, mind perception does not increase in the more friendly dialogue version.

Further research should replicate our findings using a larger sample size. Also, in a similar study the variation of friendliness in interaction had less impact on the participants' perception than the interaction context [43]. Thus, one would have to take a closer look at how politeness and context interact in future studies. In addition, related literature also suggests that anthropomorphic perceptions could be increased by increased politeness [44]. Thus, friendliness can generally be expected to have an effect on the perception of artificial agents.

The dialogue in the present study not only varied in terms of friendliness but also in terms of verbosity. It could be argued that this is not the same and a higher verbosity might have had an unwanted effect, especially on the user's task performance. Future studies could consider whether they can be designed to investigate the effect of friendliness without directly changing agent verbosity.

It would also be interesting to conduct a similar study to explore dialogue usage in the robot Biron. As he is supposed to guide the visitor to the requested room, he spends several minutes with the visitor without exchanging necessary information, thus, is can be expected that the usage of small talk affects the interaction in a positive way.

ACKNOWLEDGEMENTS

The authors would like to thank all colleagues who contributed to the Dialogue Demonstrator: Anja Durigo, Britta Wrede, Christina Unger, Christoph Broschinski, David Schlangen, Florian Lier, Frederic Siepmann, Hendrik Buschmeier, Jan De Ruiter, Johanna Müller, Julia Peltason, Lukas Twardon, Marcin Wlodarczak , Marian Pohling, Patrick Holthaus, Petra Wagner, Philipp Cimiano, Ramin Yaghoubzadeh, Sebastian Ptock, Sebastian Wrede, Thorsten Spexard, Zofia Malisz, Stefan Kopp, and Thilo Paul-Stueve. The research reported here was funded by the Cluster of Excellence "Cognitive Interaction Technology" (EXC 277). Griffiths is also partly supported by ConCreTe: the project ConCreTe acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 611733. The authors would also like to thank three anonymous reviewers of the contribution for their very useful and productive feedback.

REFERENCES

[1] Nalini Ambady, Debi LaPlante, Thai Nguyen, Robert Rosenthal, Nigel Chaumeton, and Wendy Levinson, 'Surgeons' tone of voice: A clue to malpractice history', *Surgery*, **132**(1), 5–9, (July 2002).

[2] Niklas Beuter, Thorsten P. Spexard, Ingo Lütkebohle, Julia Peltason, and Franz Kummert, 'Where is this? - Gesture based multimodal interaction with an anthropomorphic robot', in *International Conference on Humanoid Robots*. IEEE-RAS, (2008).

[3] Timothy Bickmore and Julie Cassell, 'Small talk and conversational storytelling in embodied conversational interface agents', in *Proceedings of the AAAI Fall Symposium on "Narrative Intelligence"*, pp. 87–92, (1999).

[4] Timothy Bickmore and Justine Cassell, '"how about this weather?" social dialog with embodied conversational agents', in *Proceedings of the American Association for Artificial Intelligence (AAAI) Fall Symposium on "Narrative Intelligence"*, pp. 4–8, Cape Cod, MA, (2000).

[5] Timothy Bickmore and Justine Cassell, 'Relational agents: a model and implementation of building user trust', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 396–403. ACM, (2001).

[6] John Brooke, 'SUS – a quick and dirty usability scale', *Usability evaluation in industry*, **189**, 194, (1996).

[7] Justine Cassell, 'Embodied conversational agents: representation and intelligence in user interfaces', *AI Magazine*, **22**(3), 67–83, (2001).

[8] Justine Cassell and Timothy Bickmore, 'Negotiated collusion: Modeling social language and its relationship effects in intelligent agents', *User Modeling and User-Adapted Interaction*, **13**(1-2), 89–132, (2003).

[9] Justine Cassell, Timothy Bickmore, Mark Billinghurst, Lee Campbell, Kenny Chang, Hannes Vilhjálmsson, and Hao Yan, 'Embodiment in conversational interfaces: Rea', in *Proceedings of the CHI'99 Conference*, pp. 520–527. ACM, (1999).

[10] David Crystal, *A Dictionary of Linguistics and Phonetics*, Blackwell Publishers, sixth edn., 2008.

[11] Östen Dahl, *The Growth and Maintenance of Linguistic Complexity*, John Benjamins Publishing Company, Amsterdam/Philadelphia, 2004.

[12] Nils Dahlbäck, Arne Jönsson, and Lars Ahrenberg, 'Wizard of oz studies: why and how', in *Proceedings of the 1st international conference on Intelligent user interfaces*, pp. 193–200. ACM, (1993).

[13] Friederike Eyssel and Dieta Kuchenbrandt, 'Manipulating anthropomorphic inferences about NAO: The role of situational and dispositional aspects of effectance motivation', in *2011 RO-MAN*, pp. 467–472. IEEE, (July 2011).

[14] Gernot A. Fink, 'Developing HMM-based recognizers with ESMERALDA', in *Text, Speech and Dialogue*, eds., Václav Matousek, Pavel Mautner, Jana Ocelíková, and Petr Sojka, volume 1692 of *Lecture Notes in Computer Science*, pp. 229–234. Springer Berlin Heidelberg, (1999).

[15] Susan T Fiske, Amy J C Cuddy, and Peter Glick, 'Universal dimensions of social cognition: warmth and competence.', *Trends in cognitive sciences*, **11**(2), 77–83, (February 2007).

[16] Norman M Fraser and G Nigel Gilbert, 'Simulating speech systems', *Computer Speech & Language*, **5**(1), 81–99, (1991).

[17] Dafydd Gibbon, Roger Moore, and Richard Winski, *Handbook of Standards and Resources for Spoken Language Systems*, Mouton de Gruyter, 1997.

[18] Charles Goodwin and John Heritage, 'Conversation analysis', *Annual Review of Anthropology*, **19**, 283–307, (1990).

[19] Heather M Gray, Kurt Gray, and Daniel M Wegner, 'Dimensions of mind perception.', *Science (New York, N.Y.)*, **315**(5812), 619, (February 2007).

[20] Sascha Griffiths, Ciro Natale, Ricardo Araújo, Germano Veiga, Pasquale Chiacchio, Florian Röhrbein, Stefano Chiaverini, and Reinhard Lafrenz, 'The ECHORD Project: A General Perspective', in *Gearing up and accelerating cross-fertilization between academic and industrial robotics research in Europe:*, eds., Florian Röhrbein, Germano Veiga, and Ciro Natale, volume 94 of *Springer Tracts in Advanced Robotics*, 1–24, Springer International Publishing, Cham, (2014).

[21] Axel Haasch, Sascha Hohenner, Sonja Hüwel, Marcus Kleinehagenbrock, Sebastian Lang, Ioannis Toptsis, Gernot A. Fink, Jannik Fritsch, Britta Wrede, and Gerhard Sagerer, 'Biron - the Bielefeld robot companion', in *Proc. Int. Workshop on Advances in Service Robotics*, eds., Erwin Prassler, Gisbert Lawitzky, P. Fiorini, and Martin Hägele, pp. 27–32. Fraunhofer IRB Verlag, (2004).

[22] Frank Hegel, Friederike Anne Eyssel, and Britta Wrede, 'The social robot Flobi: Key concepts of industrial design', in *Proceedings of the 19th IEEE International Symposium in Robot and Human Interactive Communication (RO-MAN 2010)*, pp. 120–125, (2010).

[23] Joseph Hilferty, 'Interview with Richard Hudson', *Bells: Barcelona English language and literature studies*, **16**, 4, (2007).

[24] Patrick Holthaus, Ingo Lütkebohle, Marc Hanheide, and Sven Wachsmuth, 'Can I help you? - A spatial attention system for a receptionist robot', in *Social Robotics*, eds., Shuzhi Sam Ge, Haizhou Li, John-John Cabibihan, and Yeow Kee Tan, pp. 325–334. IEEE, (2010).

[25] Patrick Holthaus and Sven Wachsmuth, 'Active peripersonal space for more intuitive HRI', in *International Conference on Humanoid Robots*, pp. 508–513. IEEE RAS, (2012).

[26] Patrick Holthaus and Sven Wachsmuth, 'The receptionist robot', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 329–329. ACM, (2014).

[27] Stefan Kopp, 'Social adaptation in conversational agents', *PerAda Magazine (EU Coordination Action on Pervasive Adaptation)*, (2009).

[28] Stefan Kopp and Ipke Wachsmuth, 'Synthesizing multimodal utterances for conversational agents', *Computer Animation and Virtual Worlds*, **15**(1), 39–52, (2004).

[29] Stephen C Levinson, *Pragmatics*, Cambridge University Press, Cambridge, 1983.

[30] Nikolaos Mavridis, 'A review of verbal and non-verbal human–robot in-

teractive communication', *Robotics and Autonomous Systems*, **63**, 22–35, (2015).

[31] Roger K Moore, 'From talking and listening robots to intelligent communicative machines', in *Robots that Talk and Listen – Technology and Social Impact*, 317 – 336, De Gruyter, Boston, MA, (2014).

[32] Julia Peltason, 'Position paper: Julia Peltason', in *6th Young Researchers' Roundtable on Spoken Dialogue Systems*, pp. 63 – 64, (2010).

[33] Julia Peltason, *Modeling Human-Robot-Interaction based on generic Interaction Patterns*, Ph.D. dissertation, Bielefeld University, 2014.

[34] Julia Peltason, Hannes Rieser, Sven Wachsmuth, and Britta Wrede, 'On Grounding Natural Kind Terms in Human-Robot Communication', *KI - Künstliche Intelligenz*, (March 2013).

[35] Julia Peltason and Britta Wrede, 'Modeling Human-Robot Interaction Based on Generic Interaction Patterns', in *AAAI Fall Symposium: Dialog with Robots*, pp. 80 —- 85, Arlington, VA, (2010). AAAI Press.

[36] Julia Peltason and Britta Wrede, 'Pamini: A framework for assembling mixed-initiative human-robot interaction from generic interaction patterns', in *SIGDIAL 2010: the 11th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 229–232, The University of Tokyo, (2010). Association for Computational Linguistics.

[37] Julia Peltason and Britta Wrede, 'The curious robot as a case-study for comparing dialog systems', *AI Magazine*, **32**(4), 85–99, (2011).

[38] Rajesh Ranganath, Dan Jurafsky, and Daniel A McFarland, 'Detecting friendly, flirtatious, awkward, and assertive speech in speed-dates', *Computer Speech & Language*, **27**(1), 89–115, (2013).

[39] Ehud Reiter and Robert Dale, *Building Natural Language Generation Systems*, Studies in Natural Language Processing, Cambridge University Press, Cambridge, 2000.

[40] F Röhrbein, S Griffiths, and L Voss, 'On industry-academia collaborations in robotics', Technical report, Technical Report TUM-I1338, (2013).

[41] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall International, Harlow, third int. edn., 2013.

[42] Amir Sadeghipour and Stefan Kopp, 'Embodied gesture processing: Motor-based integration of perception and action in social artificial agents', *Cognitive computation*, **3**(3), 419–435, (2011).

[43] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Effects of politeness and interaction context on perception and experience of HRI', in *Social Robotics*, eds., Guido Herrmann, Martin J. Pearson, Alexander Lenz, Paul Bremner, Adam Spiers, and Ute Leonards, volume 8239 of *Lecture Notes in Computer Science*, 531–541, Springer International Publishing, (2013).

[44] Maha Salem, Micheline Ziadee, and Majd Sakr, 'Marhaba, how may I help you?: Effects of politeness and culture on robot acceptance and anthropomorphization', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 74–81. ACM, (2014).

[45] Aaron Steinfeld, Odest Chadwicke Jenkins, and Brian Scassellati, 'The oz of wizard: simulating the human for interaction research', in *Human-Robot Interaction (HRI), 2009 4th ACM/IEEE International Conference on*, pp. 101–107. IEEE, (2009).

[46] Terry Winograd, *Language as a cognitive process (Vol. 1)*, Addison-Wesley, Reading, MA, 1983.

[47] Sebastian Wrede, Jannik Fritsch, Christian Bauckhage, and Gerhard Sagerer, 'An XML based framework for cognitive vision architectures', in *Proc. Int. Conf. on Pattern Recognition*, number 1, pp. 757–760, (2004).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Turn-yielding cues in robot-human conversation

Jef A. van Schendel and **Raymond H. Cuijpers**[1]

Abstract. If robots are to communicate with humans in a successful manner, they will need to be able to take and give turns during conversations. Effective and appropriate turn-taking and turn-yielding actions are crucial in doing so. The present study investigates the objective and subjective performance of four different turn-yielding cues performed by a NAO robot. The results show that an artificial cue, flashing eye-LEDs, lead to significantly shorter response times by the conversational partner than not giving any cue and was experienced as an improvement to the conversation. However, stopping arm movement or head turning cues showed, respectively, no significant difference or even longer response times compared to the baseline condition. Conclusions are that turn-yielding cues can lead to improved conversations, though it depends on the type of cue, and that copying human turn-yielding cues is not necessarily the best option for robots.

1 INTRODUCTION

"Beep boop!" Will our future robot partners communicate with us like Star Wars' R2D2? A more desirable future would be one where we can interact with robots in a fluent and pleasant manner, using the same natural language we use to talk to other people.

As robots grow more advanced, they are able to help us out in more areas of our lives. An area of interest is for instance elderly care, since healthcare costs in European countries are on the rise [6], and the 80+ population in Europe is expected to more than double from 2013 to 2050 [23]. Robots could increase cost-efficiency and have shown positive effects in this area [5].

But no matter what type of work, socially assistive robots as they are called [22], should be not just able to successfully perform their tasks, but deal with human beings in an appropriate, respectful and productive manner. This requires a way to naturally communicate with them, which involves taking and giving turns. This is also called managing the conversational floor.

1.1 Turn-taking

To manage the conversational floor, humans make use of turn-taking and turn-yielding cues [8]. One way to give such cues is through speech itself: the intention to yield a turn can be made clear through syntax (for instance, ending with a direct question) but also changes in intonation or speaking rate [10, 13]. Using these cues requires understanding what is being said, which is difficult for robots. Another way is through non-verbal cues, given through body movement or gaze direction [16]. The major advantage of non-verbal cues is that they do not require speech to be intelligible.

[1] Human Technology Interaction group, Eindhoven University of Technology, the Netherlands, email: r.h.cuijpers@tue.nl

Existing research has investigated ways for robots and other agents to shape and guide a conversation. Positive results have been found when robots have been used to implement conversational gaze behavior [2, 18, 21] and gestures [14, 17], likewise with agents who make use of eye gaze [1, 7, 19], especially when it is appropriate in context [9, 12]. Other researchers investigated both gestures and eye gazing by robots and, in certain combinations, found positive effects on message retention [24] and persuasion [11]. Others still moved on from dyadic sessions to conversations where a robot speaks with multiple people, so-called multiparty settings [3, 4, 15, 18, 25].

Since non-verbal cues have shown promising results in studies such as these, and can be implemented relatively easily for robots, they are of interest for the present study.

While turn-taking has been investigated in many studies, most of them evaluate a combination of turn-yielding cues as a whole and do not compare the effectiveness of isolated turn-yielding cues. Some authors, such as [4], have built interaction models for agents that include turn-yielding. In their study, the assessment of turn-yielding behavior is mixed with other types of interaction. Additionally, the subjective assessment is based on a single condition and is not compared to other models, which makes it difficult to understand the relative contribution of different turn-yielding cues. Therefore, we designed a study in which we can compare the effectiveness and user evaluation of a number of non-verbal turn-yielding cues. The response time of the conversation partner is used as an objective measure, because a shorter response time could mean better and more fluent conversational flow. Shiwa and colleagues [20] already showed that this does not necessarily signify a more pleasant interaction, which is why a questionnaire is used to evaluate the participants' opinion on the value of the different cues. This study will give us further insights in how to employ non-verbal turn-yielding and turn-taking cues during human-robot interaction.

1.2 Turn-yielding cues

Four different turn-yielding cues were selected, based on existing literature.

The first two were based on common human cues and labelled *turn head* and *stop arms*. The former means that the speaker directs its gaze away from the conversational partner during speaking, then returns to the partner when yielding the turn [16]. For the latter, the speaker uses co-speech gestures while talking, but stops doing so when finished. It is based on the idea that interlocutors make certain continuous movements during speaking, but stop moving as a sign that their turn is over [16].

For the third cue, an artificial action was chosen, namely *flash eyes*, where the robot briefly increases the brightness of its eye-LEDs. This condition was added to investigate whether cues have to be based on existing human behavior or not. This cue is not natural in the sense that it is humanlike, but it is a very common way to communicate non-verbally for robots (and many other technical devices).

The last cue was called *stay silent* and served as the baseline condition. Here, the robot simply stopped speaking with no further action.

These four cues were performed by a robot in dyadic sessions with human conversational partners. In order to generate a large number of turn-yielding events we developed a new task where the participant and the robot took turns to verbally cite the letters of the alphabet. As soon as the robot stopped citing, the participant continued citing letters. After a few letters, the robot continued again. The turn-yielding cues employed by the robot were manipulated.

2 METHOD

2.1 Participants

A total of 20 participants took part in the experiment. One was unable to complete the task and therefore the data in question was not used in the analysis. Roughly half of the participants were recruited from the J.F. Schouten participant database, while the others were recruited through word-of-mouth and invitations via social networks. The only requirement set beforehand was that the participants were able of hearing. Of the 19 participants, 13 were female. All participants were offered monetary compensation or course credits for their time.

2.2 Design

The performed experiment had a within-subjects, repeated measures design with four conditions.

The independent variable in this study was the turn-yielding cue used by the robot. The four conditions, as described under 1.2, were labelled *stay silent*, *stop arms*, *turn head* and *flash eyes*. These were randomly selected by the robot during the experiment.

The dependent variable was the response time of the participant. Specifically, this time was defined as the length in milliseconds between the start of the robot's turn-yielding cue and the beginning of the participant's speech.

Additionally, the participants filled out a questionnaire after the experiment. The questionnaire began by asking the participants which of the four cues they remember noticing. Then, a number of questions asked about their opinion on the four conditions, using a five-point Likert scale. The order of the questions was randomized for each participant in order to minimize ordering bias.

2.3 Setup

This study used a 58-centimeter tall humanoid robot called NAO, developed by Aldebaran Robotics. It has 25 degrees of freedom for movement and various sensors. Of particular interest for this study was its microphone, however, due to unsatisfactory performance during pre-tests, an external microphone was used for the experiment. Both the NAO and the microphone were connected to a laptop, used for controlling the experiment and saving the data.

The experiment took place in the GameXPLab, a laboratory modelled after a living room at Eindhoven University of Technology. Participants were seated in front of a small desk, with the NAO on top of the desk and a small wireless microphone placed between them.

2.4 Procedure

During a short introduction, the participants were given their task: together with the NAO, they were to repeatedly cite the letters of the alphabet. The NAO would start and after a randomly chosen amount of letters it would stop speaking and perform one of the turn-yielding cues. Then, the participant would continue until the NAO started speaking again. The robot autonomously decided when to speak by listening for 2, 3 or 4 utterances after which it waited for a silence to start speaking. The number of utterances determines which letter should be used next. Occasionally, the robot made a mistake (e.g. when it mistook another sound for an letter) or interrupted a person, but this was never a problem from the user's point of view. A small timing delay (0.5s) was added to make the flow as natural as possible. This cycle continued for roughly 15 minutes with each participant.

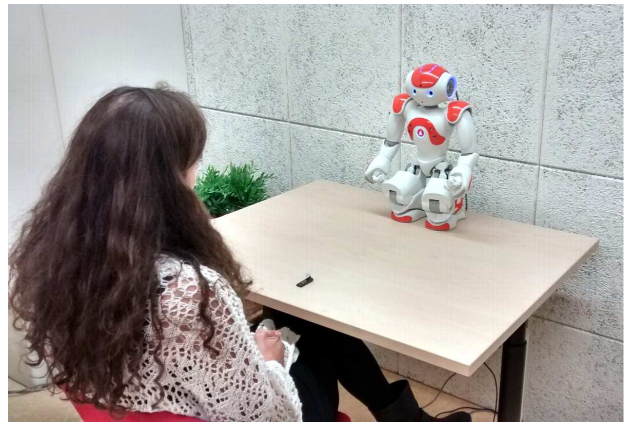


Figure 1. Experiment setup

This particular task was chosen for several reasons. First, the answers by the participants would mostly be single-syllable words, which would make them easier to accurately detect with the microphone and enable the robot to count them, so it would know where to continue the series. The second reason was the assumption that the participants would be able to recall the letters of the alphabet with minimal effort, thereby minimizing the influence of recollection time. Thirdly, the advantage of using a fixed sequence would be to avoid the need for the participant to decide on what to say. In other words, the aim was to control for possibly confounding variables such as recollection time or deliberation time.

Afterwards, the participants filled out a questionnaire (further described under 2.2).

3 RESULTS

3.1 Experiment results

The experiment data was edited and analyzed using SPSS. A number of false positives were recorded as notes during the experiment. After these were removed, a total of 1310 valid data points were left, or about 68.9 recorded measurements per participant.

The distribution of the response time data was found to be skewed right ($\text{skewness} = 1.520 \pm 0.068$) and peaked ($\text{kurtosis} = 5.370 \pm 0.135$). To increase normality it was logarithmically transformed. Histograms of the original (a) and log-transformed (b) data can be found in Figure 3.1. As can be seen, the normality was much improved: the distribution of the transformed data is approximately symmetric ($\text{skewness} = -0.079 \pm 0.068$) and less peaked (kurtosis $= 0.421 \pm 0.135$).

Table 1 shows the reaction times of the four conditions. Since the distribution of reaction times is skewed we transformed the data using the natural logarithm (ln) before computing the means and standard errors (middle two columns). The last two columns show the reaction times transformed back to the normal time domain.

A one-way ANOVA showed that there was a significant difference between groups ($F(3, 1306) = 15.407$, $p < 0.001$). Levene's test indicated equal variances ($p = 0.644$).

A Tukey HSD post-hoc test revealed that the response time was significantly lower for the *flash eyes* action ($M = 854$ ms, $p = 0.006$) yet significantly higher for the *turn head* action ($M = 1033$ ms, $p = 0.003$) when compared to the *stay silent* condition (M = 944 ms). There was no significant difference between the *stay silent* condition and the *stop arms* action ($M = 916$ ms, $p = .829$).

Additionally, the mean response time for the *turn head* condition was significantly higher than both the *stop arms* ($p < 0.001$) and *flash eyes* ($p < 0.001$) conditions. There was, however, no significant difference between the *flash eyes* and *stop arms* conditions ($p = 0.071$). Post-hoc results are shown in Table 2. A bar chart visualising the means of the four conditions can be found in Figure 4.

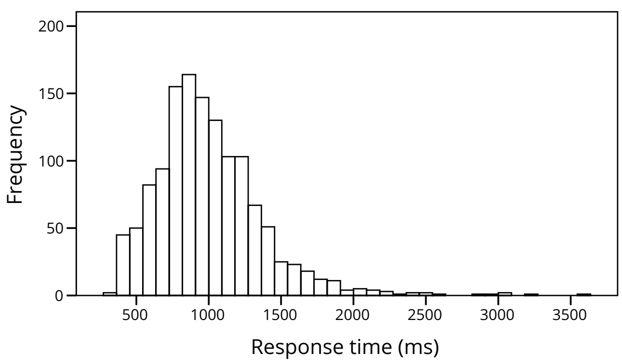


Figure 2. Original data

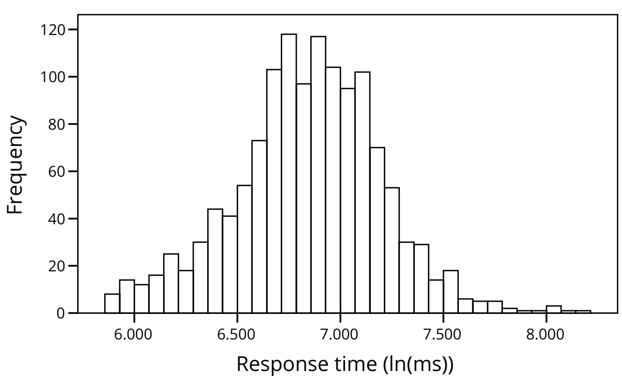


Figure 3. Histograms showing the distribution of response times

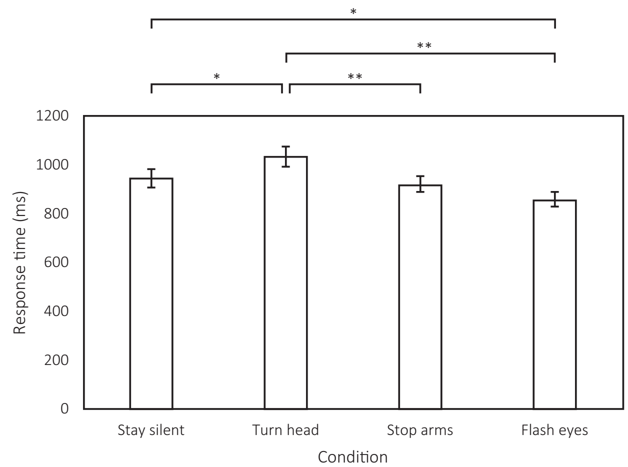


Figure 4. Means of the four conditions. Error bars represent 95% CI. Bars denoted with * differ at significance level < 0.01, bars with ** at significance level < 0.001.

Linear regression on the response times with trial number as the independent variable showed that these times did not decrease after sequential trials (*stay silent* $p = 0.759$; *turn head* $p = 0.224$; *flash eyes* $p = 0.368$), except for the *stop arms* condition ($p = 0.001$). For this last condition, response times decreased by 207 ms after 115 trials, as shown in Figure 5.

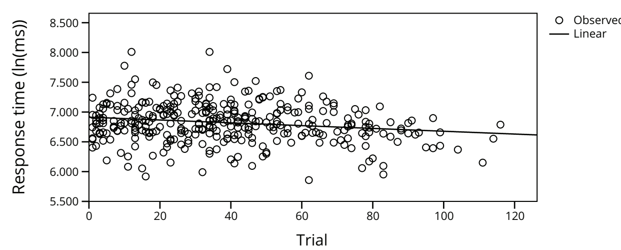


Figure 5. Scatter plot and fitted line of all response times in the *stop arms* condition.

3.2 Questionnaire results

The data gathered with the questionnaire (N = 19) was edited and analyzed using SPSS, in several steps.

The first part of the questionnaire was used as a confirmation of which cues were noticed by the participants. Cues that went unnoticed were excluded from the data.

Furthermore, the questionnaire included pairs of opposite questions, phrased positively and negatively, to avoid acquiescent bias. An example of such a pair is "...improved the flow of the conversation" and "...did not improve the conversation". Before analysis, negatively phrased questions had their answers mirrored.

Principle component analysis was used to identify the underlying factors and group the variables. After applying varimax rotation,

Table 1. Reaction times of the four conditions in the log-transformed and normal domain. SE is the standard error of sample mean. N is the number of turn yields (1310 in total).

Condition	N	Mean (ln(ms))	SE (ln(ms))	Mean (ms)	SE (ms)
Stay silent	331	6.85	.020	944	$\pm$19
Turn head	337	6.94	.019	1033	20/-19
Stop arms	334	6.82	.018	916	17/-16
Flash eyes	308	6.75	.020	854	$\pm$17

Table 2. Post-hoc test results of the response times

(I) condition	(J) condition	Mean difference (I-J, ln(ms))	SE (ln(ms))	Sig.
Stay silent	Turn head	-0.95	.027	.003
	Stop arms	.023	.027	.829
	Flash eyes	.091	.028	.006
Turn head	Stay silent	.095	.027	.003
	Stop arms	.118	.027	.000
	Flash eyes	.186	.028	.000
Stop arms	Stay silent	-.023	.027	.829
	Turn head	-.118	.027	.000
	Flash eyes	.068	.028	.071
Flash eyes	Stay silent	-.091	.028	.006
	Turn head	-.186	.028	.000
	Stop arms	-.068	.028	.071

three components were found with an Eigenvalue over 1, accounting for 35.1, 28.2 and 13.2 percent, respectively, of the total variance.

The rotated component matrix, shown in Table 3, shows which questions load on which components after rotation. Based on this data, the three components were named *Pleasant*, *Improvement* and *Noticeable*. Table 4 shows which questions make up which components.

After identifying the components, a one-way ANOVA on the combined questions showed that there was a significant difference between groups for the *Improvement* ($F(3, 292) = 8.998$, $p < 0.001$) and *Noticeable* ($F(3, 70) = 3.081$, $p = 0.033$) components, but not for the *Pleasant* component ($F(3, 218) = 0.602$, $p = 0.614$).

A Tukey HSD post-hoc test performed on the *Improvement* and *Noticeable* components showed that there were several significant differences between the means of the questionnaire responses. *Flash eyes* scored significantly higher on *Improvement* than both *stop arms* ($p < 0.001$) and *stay silent* ($p = 0.001$). Also, *stop arms* scored higher than *stay silent* on *Noticeable* ($p = 0.040$).

The post-hoc test results for the *Improvement* and *Noticeable* components can be found in Table 5 and 6, respectively. A graphical summary of all the components can be found in Figure 6.

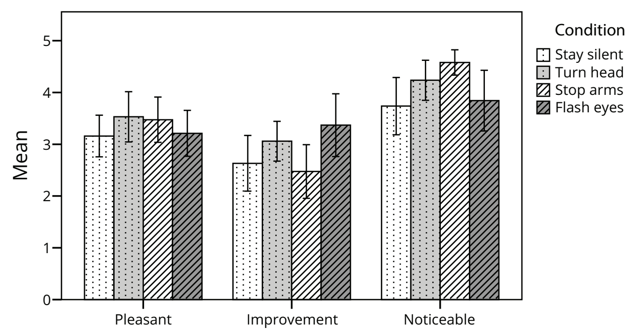


Figure 6. Means of the four conditions for every component. Error bars represent 95% CI.

4 DISCUSSION

The present study investigated different turn-yielding cues to be used by a robot in robot-human conversation. An experiment and questionnaire measured the performance and rating of the different cues. The results show that using a turn-yielding cue can lead to faster response times by the conversational partner compared to the baseline condition. One of the cues, namely *flash eyes*, produced the lowest response times and was rated higher on Improvement than the baseline condition and any other cue. The results, therefore, partially confirm the hypothesis that turn-yielding cues by a robot can improve robot-human conversation.

4.1 Different types of cues

The *flash eyes* cue lead to faster response times and had the highest *Improvement* rating by the participants. However, other cues showed different results. The *turn head* cue showed significantly longer response times compared to *staying silent*. Moreover, while the *stop arms* condition was rated as more noticeable than *staying silent*, there was no significant difference between the mean response times of these two cues.

There was a difference of 179 ms between the means of the response times for the *flash eyes* and *turn head* cues. A conclusion could be that while turn-yielding cues have the potential to lead to decreased response times, the type of cue matters a great deal.

Table 3. Rotated component matrix. Questions marked with * were mirrored.

Question	Component 1	Component 2	Component 3
...made it obvious it was my turn	.913	.094	.035
...had no clear meaning*	.868	.023	-.110
...did not improve the conversation*	.723	.459	.067
...improved the flow of the conversation	.703	.465	-.143
...was uncomfortable*	.074	.871	-.101
...was friendly	.142	.863	.155
...felt natural	.415	.560	-.096
...was hard to notice*	-.060	-.007	.986

Table 4. Components and related questions. Questions marked with * were mirrored.

Component 1, *Pleasant*	Component 2, *Improvement*	Component 3, *Noticeable*
...was uncomfortable*	...made it obvious it was my turn	...was hard to notice*
...was friendly	...had no clear meaning*	
...felt natural	...did not improve the conversation*	
	...improved the flow of the conversation	

4.2 Artificial cue

While a decrease in response time can be a hint that the cue improves the conversation, this does not necessarily have to be the case. Results from the questionnaire, however, were in line with the results from the experiment when it came to the *flash eyes* cue. It was seen as an improvement to the conversation and to have a clearer meaning when compared to the *stop arms* and *stay silent* cues.

Some anecdotal evidence from the experiment pointed the same way. Several participants remarked that they appreciated the *flash eyes* cue, one of them explaining "It signals that he is done, and that he won't interrupt me". Multiple participants also described the cue as "natural", which is interesting for an artificial cue that human conversational partners are unable to perform.

Thus, one of the interesting things here is that the cue with the lowest response time was an artificial cue, as opposed to the *turn head* and *stop arms* cues, which were based on literature from human-human interaction. There appears to be a difference between a human being using such cues and the NAO doing the same. This could have several causes. One possible cause is that the NAO did not perform the cue correctly, and therefore its meaning was unclear to the participants. Results from the questionnaire are inconclusive on this point: these cues were not rated significantly lower on this point, and their means center around "Neither agree nor disagree". Another reason could be that the participants found the cues with movement to be unexpected and therefore hesitated in their responses.

4.3 Movement cues

The cues that were based on movement, namely *turn head* and *stop arms*, showed worse performance compared to *flash eyes*, which did not involve movement. The movements made by the robot could be a source of distraction or hesitation for the participants, which could explain the longer response times.

Some anecdotal evidence from the experiment pointed this way. Some of the participants talked about the *turn head* and *stop arms* cues, explaining that they found many of the robot's movements to be distracting, and were sometimes confused as to the meaning of these movements. The data from the questionnaire shows that the *stop arms* cue was rated as significantly higher on the *Noticeable* component. Could it have been too noticeable, thereby distracting the participant?

Additionally, during the experiment it often seemed that when the NAO started moving, the participant hesitated to continue, preferring to wait to see where the robot was going with this. One of them remarked that he did not recognize the movement of *turn head* as a cue to start speaking, so instead he "just waited until it was done".

The movements could have simply been unexpected. Linear regression showed that for at least the *stop arms* cue, the mean response time decreased after subsequent trials, suggesting the participants were faster to respond and perhaps got used to the cue. Perhaps after longer interaction with the robot, this cue could have lead to response times similar to *flash eyes*.

Whether these findings are specific to the NAO robot is unclear, but fact is that this particular robot makes distinct sounds during movements and that it remains completely static outside of the performed cues. This could make movement cues highly salient by default.

4.4 Improvements to the experiment

A critical component of the experiment was accurately measuring the response time. The external microphone made it possible to relatively accurately and precisely measure the points at which the participant started speaking. However the beginning of the measurement, defined as the point at which the NAO stopped speaking, was harder to measure accurately. In the experiment, the timer started running after the NAO signalled it was done. However further investigation revealed that there is in fact a pause between the actual end of the sound and this signal, of around 225 ms on average. Though this issue could unfortunately not be avoided during this experiment, it could have an impact on the results. In practice it means that the turn-yielding cue could be performed sooner after speaking, possibly leading to a larger decrease in response times and an even stronger effect. Indeed, if we subtract 225ms form the reaction times for all non-verbal cues except the *stay silent* cue in Figure 4, we obtain a graph where all non-verbal cues lead to a reaction time improvement

Table 5. Post-hoc test results for the *Improvement* component

(I) condition	(J) condition	Mean difference (I-J)	SE	Sig.
Flash eyes	Turn head	.449	.193	.096
	Stop arms	.921	.188	.000
	Stay silent	.724	.188	.001
Turn head	Flash eyes	-.449	.193	.096
	Stop arms	.472	.193	.072
	Stay silent	.275	.193	.488
Stop arms	Flash eyes	-.921	.188	.000
	Turn head	-.472	.193	.072
	Stay silent	-.197	.188	.720
Stay silent	Flash eyes	-.724	.188	.001
	Turn head	-.275	.193	.488
	Stop arms	.197	.188	.720

Table 6. Post-hoc test results for the *Noticeable* component

(I) condition	(J) condition	Mean difference (I-J)	SE	Sig.
Flash eyes	Turn head	-.393	.319	.608
	Stop arms	-.737	.310	.091
(p ¡ .001)	Stay silent	.105	.310	.986
Turn head	Flash eyes	.393	.319	.608
	Stop arms	-.344	.319	.704
	Stay silent	.498	.319	.406
Stop arms	Flash eyes	.737	.310	.091
	Turn head	.344	.319	.704
	Stay silent	.842	.310	.040
Stay silent	Flash eyes	-.105	.310	.986
	Turn head	-.498	.319	.406
	Stop arms	-.842	.310	.040

compared to the *stay silent* cue. However, the *flash eyes* cue would still be most salient and the relative effectiveness of these cues remains the same.

5 CONCLUSIONS

The present study explored the use of turn-yielding cues by a robot. We found that such turn-yielding cues can improve both performance and user experience during human-robot conversation. These results on turn-yielding are in line with earlier findings that show that non-verbal cues can influence turn taking in conversations [2, 18]. Our study adds to earlier research by specifically focusing on the relative effect of turn-yielding cues and it shows that the type of cue is of importance for both performance and user experience.

An important question is how these conclusions are to be used in the development of socially assistive robots. Should one, for instance, always make use of an eye-flashing cue? It is clear that turn-yielding cues have the potential to improve a conversation, but in our study at most one cue was presented at a time (in addition to the stay silent cue). While the eye-flashing cue showed the most promise during this experiment, its meaning is, in general, ambiguous. Flashing LEDs are used to signal all sorts of events. In that sense the *turn head* and *stop arms* cues are much better, because they not only inform the observer about the timing of an event but also that the event is a turn-yield. So we expect that these cues are more useful in complex interactions. Finally, it would be interesting to see how these cues interact. A head turn could disambiguate a LED flash, so that in combination the turn-yield cues are effective and robust.

REFERENCES

[1] Sean Andrist, Tomislav Pejsa, Bilge Mutlu, and Michael Gleicher, 'Designing effective gaze mechanisms for virtual agents', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 705–714, (2012).

[2] Sean Andrist, Xiang Zhi Tan, Michael Gleicher, and Bilge Mutlu, 'Conversational gaze aversion for humanlike robots', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pp. 25–32, (2014).

[3] Maren Bennewitz, Felix Faber, Dominik Joho, Michael Schreiber, and Sven Behnke, 'Integrating vision and speech for conversations with multiple persons', in *International Conference on Intelligent Robots and Systems 2005*, pp. 2523–2528, (2005).

[4] Dan Bohus and Eric Horvitz, 'Multiparty turn taking in situated dialog: Study, lessons, and directions', in *Proceedings of the SIGDIAL 2011 Conference*, pp. 98–109, (2011).

[5] Joost Broekens, Marcel Heerink, and Henk Rosendal, 'Assistive social robots in elderly care: a review', *Gerontechnology*, **8**(2), 94–103, (2009).

[6] Sarah Chaytor and Uta Staiger, 'The future of healthcare in europe', *UCL European Institute*, (2011).

[7] Alex Colburn, Michael F. Cohen, and Steven Drucker, 'The role of eye

gaze in avatar mediated conversational interfaces', *Sketches and Applications, Siggraph'00*, (2000).

[8] Starkey Duncan, 'Some signals and rules for taking speaking turns in conversations.', *Journal of personality and social psychology*, **23**(2), 283, (1972).

[9] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse, 'The impact of eye gaze on communication using humanoid avatars', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 309–316, (2001).

[10] Agustín Gravano and Julia Hirschberg, 'Turn-yielding cues in task-oriented dialogue', in *Proceedings of the SIGDIAL 2009 Conference: The 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pp. 253–261, (2009).

[11] Jaap Ham, René Bokhorst, Raymond Cuijpers, David van der Pol, and John-John Cabibihan, 'Making robots persuasive: the influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power', in *Social Robotics*, 71–83, Springer, (2011).

[12] D. K. J. Heylen, Ivo Van Es, Anton Nijholt, and E. M. A. G. van Dijk, 'Experimenting with the gaze of a conversational agent', (2002).

[13] Anna Hjalmarsson, 'The additive effect of turn-taking cues in human and synthetic voice', *Speech Communication*, **53**(1), 23–35, (2011).

[14] Chien-Ming Huang and Bilge Mutlu, 'Modeling and evaluating narrative gestures for humanlike robots.', in *Robotics: Science and Systems*, (2013).

[15] Martin Johansson, Gabriel Skantze, and Joakim Gustafson, 'Head pose patterns in multiparty human-robot team-building interactions', in *Social Robotics*, 351–360, Springer, (2013).

[16] Adam Kendon, 'Some functions of gaze-direction in social interaction', *Acta psychologica*, **26**, 22–63, (1967).

[17] Chaoran Liu, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita, 'Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction', in *Human-Robot Interaction (HRI), 2012 7th ACM/IEEE International Conference on*, pp. 285–292, (2012).

[18] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita, 'Footing in human-robot conversations: how robots might shape participant roles using gaze cues', in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 61–68, (2009).

[19] David G. Novick, Brian Hansen, and Karen Ward, 'Coordinating turn-taking with gaze', in *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pp. 1888–1891, (1996).

[20] T. Shiwa, T. Kanda, M. Imai, H. Ishiguro, and N. Hagita, 'How quickly should communication robots respond?', in *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pp. 153–160, (March 2008).

[21] Candace L. Sidner, Cory D. Kidd, Christopher Lee, and Neal Lesh, 'Where to look: a study of human-robot engagement', in *Proceedings of the 9th international conference on Intelligent user interfaces*, pp. 78–84, (2004).

[22] Adriana Tapus, Mataric Maja, and Brian Scassellatti, 'The grand challenges in socially assistive robotics', *IEEE Robotics and Automation Magazine*, **14**(1), (2007).

[23] UN, 'World population prospects, the 2012 revision', *New York: Department for Economic and Social Affairs*, (2013).

[24] Elisabeth T. Van Dijk, Elena Torta, and Raymond H. Cuijpers, 'Effects of eye contact and iconic gestures on message retention in human-robot interaction', *International Journal of Social Robotics*, **5**(4), 491–501, (2013).

[25] Roel Vertegaal, Robert Slagter, Gerrit Van der Veer, and Anton Nijholt, 'Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 301–308, (2001).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Robot Learning from Verbal Interaction: A Brief Survey

Heriberto Cuayáhuitl[1]

Abstract. This survey paper highlights some advances and challenges in robots that learn to carry out tasks from verbal interaction with humans, possibly combined with physical manipulation of their environment. We first describe what robots have learnt from verbal interaction, and how do they do it. We then enumerate a list of research limitations to motivate future work in this challenging and exciting multidisciplinary area. This brief survey points out the need of bringing robots out of the lab, into uncontrolled conditions, in order to investigate their usability and acceptance by end users.

1 INTRODUCTION

Intelligent conversational robots are an exciting and important area of research because of their potential to provide a natural language interface between robots and their end users. A learning conversational robot can be defined as an entity which improves its performance over time through verbally interacting with humans and/or other machines in order to carry out abstract or physical tasks in its (real or virtual) world. The vision of such kinds of robots is becoming more realistic with technological advances in artificial intelligence and robotics. The increasing development of robot skills presents boundless opportunities for them to perform useful tasks for and with humans. Such development is well suited to robots with a physical body because they can exploit their input and output modalities to deal with the complexity of public spatial environments such as homes, shops, airports, hospitals, etc. A robot learning from interaction, rather than a robot that does not learn, is particularly relevant because it is not feasible to pre-program robots for all possible environments, users and tasks. Even though many robotic systems can be scripted or programmed to behave just as expected, the rich nature of interaction with the physical world, or with humans, demands flexible, adaptive solutions to deal with dynamic, previously unknown, or highly stochastic domains. Therefore, robots should be able to refine their already learned skills over time and/or acquire new skills by (verbally) interacting with its users and its spatial environment. An emerging multidisciplinary community at the intersection of machine learning, human-robot interaction, natural language processing, robot perception, robot manipulation and robot gesture generation, among others, seeks to address challenges in realising such robots capable of interactive learning.

This paper will provide a brief survey on robots that learn to acquire or refine their verbal skills from example interactions using machine learning. Conversational robots that draw on hand-coded behaviours, or robots learning from non-verbal interaction [3, 14], are therefore considered out of scope here.

[1] Heriot-Watt University, United Kingdom, email: hc213@hw.ac.uk

2 ADVANCES

2.1 What have robots learnt from conversational interaction?

The following list of representative conversational robots shows a growing interest in this multidisciplinary field, see Figure 1.

- The mobile robot *Florence* is a nursing home assistant [20, 17]. The tasks of this robot include providing the time, providing information about the patient's medication schedule and TV channels, and motion commands such as go to the kitchen/bedroom. The learning task consists in inducing a dialogue strategy under uncertainty, where the actions correspond to physical actions (motion commands) and clarification or confirmation actions. The robot's goal is to choose as many correct actions as possible.
- Iwahashi's non-mobile robot with integrated arm+hand+head learns to communicate from scratch by physically manipulating objects on a table [11]. The tasks of this robot include (a) acquisition of words, concepts and grammars for objects and motions; (b) acquisition of the relationships between objects; and (c) the ability to answer questions based on beliefs. The robot's goal is to understand utterances and to generate reasonable responses from a relatively small number of interactions.
- The mobile robot *SmartWheeler* is a semi-autonomous wheelchair for assisting people with severe mobility impairments [19]. The task of the robot is to assist patients in their daily locomotion. The learning task is similar as in the Florence robot, the induction of a dialogue manager under uncertainty, but with a larger state space (situations). The robot's goal is to reduce the physical and cognitive load required for its operation.
- A mobile robotic forklift is a prototype for moving heavy objects from one location to another [25]. Example commands include going to locations, motion commands, and picking up and putting down objects. The learning task consists in understanding natural language commands in the navigation and object manipulation domain. The robot's goal is to ground natural language commands (mapping commands to events, objects and places in the world [18]) in order to output a plan of action.
- The humanoid robot *Simon* manipulates physical objects on a table from human teachers [2]. The task of the robot includes pouring cereal into bowls, adding salt to salads, and pouring drinks into cups. The learning task is to ask questions to human demonstrators from three different types: label queries (Can I do it like this?), demonstration queries (Can you show me how to do it?) , and feature queries (Should I keep this orientation?). The robot's goal is to ask as good questions as possible in order to achieve fast learning from physical demonstrations.
- A KUKA mobile platform with manipulator ensembles simple furniture [24]. The task of the robot is to assemble IKEA furniture such as tables based on STRIPS-like commands. The learning

Figure 1. Example learning conversational robots: (a) Florence nursebot [20], (b) Iwahashi's robot [11], (c) Kuka furniture assembler [24], (d) Nao giving directions [1], (e) Nao playing quizzes [7], (f) Simon robot learning from demonstrations [2], (g) James bartender robot [12], (h) Forklift robot [25], (i) SmartWheeler [19], (j) PR2 learning new words [15], (k) Gambit picking up objects [16], and (l) Cosero receiving verbal commands [21]. See text in Section 2.

tasks consists in learning to ground language and to train a natural language generator in order to ask for help to humans (by generating words from symbolic requests) when the robot encounters a failure situation. The robot's goal is to ensemble furniture as independently as possible and to ask for help when failures occurred.

- The torso robot *James* serves drinks to people in a pub [12]. The task of the robot is to approach customers in natural language, to ask for the drinks they want, and to serve the requested drinks. The learning task consists in inducing a dialogue manager for multi-party interaction. The robot's goal is to serve as correct drinks as possible based on socially acceptable behaviour due to the presence of multiple customers at once in the robot's view.
- The humanoid robot *NAO* has been used to play interactive quiz games [7, 6]. The robot's tasks include engaging into interactions, asking and answering questions from different fields, and showing affective gestures aligned with verbal actions. The learning task consists in inducing a dialogue strategy optimising confirmations and flexible behaviour, where users are allowed to navigate flexibly across subdialogues rather than using a rigid dialogue flow. The robot's goal is to answer correctly as much as possible and to ask as many questions as possible from a database of questions.
- The humanoid robot *NAO* has been used to give indoor route instructions [1]. The task of the robot is to provide directions, verbally and with gestures, to places within a building such as offices, conference rooms, kitchen, cafeteria, bathroom, etc., based on a predefined map. The learning task is to induce a model of engagement to determine when to engage, maintain or disengage an interaction with the person(s) in front of the robot. The robot's goal is to direct people to the locations they are looking for.
- The mobile robot *PR2* has been used to acquire new knowledge of objects and their properties [15]. The tasks of the robot include to spot unknown objects, to ask how unknown objects look like, and to confirm newly acquired knowledge. The learning task is to extend its knowledge base of objects via descriptions of their physical appearance provided by human teachers. The robot's goal is to answer questions of its partially known environment.
- The robot arm *Gambit* has been used to study how users users refer to groups of objects with speech and gestures. The tasks of the robot is to move indicated objects in a workspace, via verbal descriptions of object properties and possibly including gestures. The learning task is to understand user intentions without requiring specialized user training. The robot's goal is to select, as correctly as possible, the referred objects on the table.
- The mobile robot *Cosero* has been used in the RoboCup at home competition, which has won several of them in recent years [21]. The tasks of the robot include to safely follow a person, to detect an emergency from a person calling for help, to get to know and recognise people and serve them drinks, and to bring objects from one location to another. The learning task is to extend its knowledge of locations, objects and people. The robot's goal is to carry out tasks autonomously—provided in spoken language—as expected and in a reasonable amount of time.

ID	Dimension / Reference	[20]	[11]	[19]	[25]	[2]	[24]	[12]	[7]	[1]	[15]	[16]	[21]	ALL
01	Learning To Interpret Commands	1	1	1	1	1	1	1	1	1	1	1	1	12
02	Dialogue Policy Learning	1	0	1	0	1	0	1	1	0	0	0	0	5
03	Learning To Generate Commands	0	1	0	0	0	1	0	0	0	0	0	0	2
04	Learning To Engage	0	0	0	0	0	0	1	1	1	0	0	0	3
05	Grammar Learning	0	1	0	0	0	0	0	0	0	0	0	0	1
06	Flexible Interaction	0	0	0	0	0	0	0	1	0	0	1	1	3
07	Speech-Based Perception	1	1	1	0	1	0	1	1	1	1	1	1	10
08	Language Grounding	0	1	0	0	0	1	0	0	0	0	1	0	3
09	Speech Production	1	1	1	0	1	0	1	1	1	1	0	1	9
10	Multimodal Fussion	0	1	1	0	1	0	1	0	1	1	1	1	8
11	Multimodal Fission	0	1	1	0	0	0	1	1	1	0	0	1	6
12	Multiparty Interaction	0	0	0	0	0	0	1	0	1	0	0	0	2
13	Route Instruction Giving	0	0	0	0	0	0	0	0	1	0	0	0	1
14	Navigation Commands	1	0	1	1	0	1	0	0	0	0	0	1	5
15	Object Recognition and Tracking	0	1	0	1	1	1	1	0	0	1	1	1	8
16	Human Activity Recognition	0	0	0	0	0	0	0	0	1	0	0	1	2
17	Localisation and Mapping	1	0	1	1	0	0	0	0	0	0	0	1	4
18	Gesture Generation	0	0	0	0	1	0	0	1	1	1	0	1	5
19	Object Manipulation	0	1	0	1	1	1	1	0	0	0	1	1	7
20	Supervised Learning	0	1	0	1	0	1	1	1	1	0	1	0	7
21	Unsupervised Learning	0	0	1	0	0	0	0	0	0	0	1	0	2
22	Reinforcement Learning	1	0	1	0	0	0	1	1	0	0	0	0	4
23	Active Learning	0	0	0	0	1	0	0	0	0	0	0	0	1
24	Learning From Demonstration	0	0	0	0	1	0	0	0	0	1	0	1	3
25	Evaluation w/Recruited Participants	1	0	0	0	1	1	1	1	1	0	1	1	8
26	Evaluation in Noisy/Crowded Spaces	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 1. Features of robots acquiring/using their verbal skills. While boolean values are rough indicators, real values are better indicators but harder to obtain.

2.2 How do conversational robots learn to interact?

Machine learning frameworks are typically used to equip robots with learning skills, and they differ in the way they treat data and the way they process feedback [13, 8]. Some machine learning frameworks addressed by previous related works are briefly described as follows:

- *Supervised learning* can be used whenever it comes to the task of classifying and predicting data, where the data consists of labelled instances (pairs of features and class labels). The task here is to induce a function that maps the unlabelled instances to labels. This function is known as a classifier when the labels are discrete and as a regressor when the labels are continuous. Conversational robots make use of classifiers to predict spatial description clauses [25], grounded language [11, 24], social states [12], dialogue acts [7], gestures [16], and engagement actions [1], among others.
- *Reinforcement Learning* makes use of indirect feedback typically based on numerical rewards given during the interaction, and the goal is to maximise the rewards in the long run. The environment of a reinforcement learning agent is represented with a Markov Decision Process (MDP) or a generalisation of it. Its solution is a policy that represents a weighted mapping from states (situations that describe the world) to verbal and/or physical actions, and can be found through a trial and error search in which the agent explores different action strategies in order to select the one with the highest payoff. This framework can be seen as a very weak form of supervised learning, where the impact of actions is rated according to the overall goal (e.g. fetching and delivering an object or playing a game). This form of learning has been applied to design the dialogue strategies of interactive robots using MDPs [12], Semi-MDP to scale up to larger domains [7], and Partially Observable MDPs to address interaction under uncertainty [20, 19].
- *Unsupervised learning* addresses the challenge of learning from unlabelled data. Since it does not receive any form of feedback, it has to find patterns in the data solely based on its observable features. The task of an unsupervised learning algorithm is thus to uncover hidden structure in unlabelled data. This form of machine learning has been used by [19] to cluster the observation space of a POMDP-based dialogue manager, by [12] to cluster social states for multiparty interaction, and by [16] to select features for gesture recognition tasks.
- *Active learning* includes a human directly within the learning procedure assuming three data sets: a small set of labelled examples, a large set of unlabelled examples, and chosen examples. The latter are built in an interactive fashion by an active learning algorithm who queries a human annotator for labels it is most uncertain of. This form of learning has been applied to *learning from demonstration* scenarios by [2] and closely related by [15, 21].

Other forms of machine learning that can be applied to conversational robots include transfer and multi-task learning, lifelong learning, and multiagent learning, among others [8, 4]. Furthermore, while a single form of learning can be incorporated into conversational robots, combining multiple forms of machine learning can be used to address perception, action and communication in a unified way. The next section describes some challenges that require further research for the advancement of intelligent conversational robots.

3 Challenges: What is missing?

Table 1 shows a list of binary features for the robots described above. These features are grouped according to language, robotics, learning, and evaluation. The lowest numbers in the last column indicate the dimensions that have received little attention. From this table, it can observed that the main demand to be addressed is conversational robots that interact with real people in uncontrolled environments rather than recruited participants in the lab. The research directions demanding further attention are briefly described as follows:

- **Noise and crowds:** most (if not all) interactive robots have been trained and tested in lab or controlled conditions, where no noise or low levels of noise are exhibited–see Table 1. A future direction concerning the whole multidisciplinary community lies in training and evaluating interactive robots in environments including people with real needs. This entails dealing with dynamic and varying levels of noise (from low to high), crowded environments on the move, distant speech recognition and understanding [26, 23] possibly combined with other modalities [5], and real users from the general population rather than just recruited participants.
- **Unknown words and meanings:** most interactive robots have been equipped with static vocabularies and lack grammar learning (see line 5 in Table 1), where the presence of unseen words lead to misunderstandings. Equipping robots with mechanisms to deal with the unknown could potentially make them more usable in the real world. This not only involves language understanding but also language generation applied to situated domains [9].
- **Fluent and flexible interaction:** when a robot is equipped with verbal skills, it typically uses a rigid turn-taking strategy and a predefined dialogue flow (see line 6 in Table 1). Equipping robots with more flexible turn-taking and dialogue strategies, so that people can say or do anything at any time, would contribute towards more fluent and natural interactions with humans [7].
- **Common sense spatial awareness:** most conversational robots have been equipped with little awareness of the dynamic entities and their relationships in the physical world (see lines 13 and 16 in Table 1). When a robot is deployed in the wild, it should be equipped with basic spatial skills to plan its verbal and non-verbal behaviour. In this way, spatial representations and reasoning skills may not only contribute to safe human-robot interactions but also with opportunities to exhibit more socially-acceptable behaviour. See [22, 10] for detailed surveys on social interactive robots.
- **Effective and efficient learning from interaction:** interactive robots are typically trained in simulated or controlled conditions. If a robot is to interact in the wild, it should be trained with such kinds of data. Unfortunately, that is not enough because moving beyond controlled conditions opens up multiple challenges in the way we train interactive robots such as the following:
  - robot learning from unlabelled or partially labelled multimodal data (see lines 21 and 23 in Table 1) should produce safe and reasonable behaviours;
  - altering the robot's behaviour, even slightly, should be straightforward rather than requiring a substantial amount of human intervention (e.g. programming);
  - inducing robot behaviours should exploit past experiences from other domains rather than inducing them from scratch; and
  - learning to be usable and/or accepted by people from the general population is perhaps the biggest challenge.

4 Conclusion

Previous work has shown the increase in multidisciplinary work to realise intelligent conversational robots. Although several challenges remain to be addressed by specialised communities, addressing them as a whole is the end-to-end challenge that sooner or later it has to be faced. This challenge involves two crucial actions with little attention so far (a) to bring robots out of the lab to public environments, and (b) to demonstrate that they are usable and accepted by people from the general public. We hope that the topics above will encourage further multidisciplinary discussions and collaborations.

REFERENCES

[1] Dan Bohus, Chit W. Saw, and Eric Horvitz, 'Directions robot: In-the-wild experiences and lessons learned', in *AAMAS*, (2014).

[2] Maya Cakmak and Andrea Lockerd Thomaz, 'Designing robot learners that ask good questions', in *HRI*, (2012).

[3] Sonia Chernova and Andrea Lockerd Thomaz, *Robot Learning from Human Teachers*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2014.

[4] Heriberto Cuayáhuitl and Nina Dethlefs, 'Dialogue systems using online learning: Beyond empirical methods', in *NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Comm.*, (2012).

[5] Heriberto Cuayáhuitl, Lutz Frommberger, Nina Dethlefs, Antoine Raux, Matthew Marge, and Hendrik Zender, 'Introduction to the special issue on machine learning for multiple modalities in interactive systems and robots', *TiiS*, **4**(3), (2014).

[6] Heriberto Cuayáhuitl and Ivana Kruijff-Korbayová, 'An interactive humanoid robot exhibiting flexible sub-dialogues', in *NAACL-HLT*, (2012).

[7] Heriberto Cuayáhuitl, Ivana Kruijff-Korbayová, and Nina Dethlefs, 'Nonstrict hierarchical reinforcement learning for interactive systems and robots', *TiiS*, **4**(3), 15, (2014).

[8] Heriberto Cuayáhuitl, Martijn van Otterlo, Nina Dethlefs, and Lutz Frommberger, 'Machine learning for interactive systems and robots: A brief introduction', in *MLIS*. ACM ICPS, (2013).

[9] Nina Dethlefs and Heriberto Cuayáhuitl, 'Hierarchical reinforcement learning for situated natural language generation', *Natural Language Engineering*, **FirstView**, (12 2014).

[10] Terrence Fong, Illah R. Nourbakhsh, and Kerstin Dautenhahn, 'A survey of socially interactive robots', *Robotics and Autonomous Systems*, **42**(3-4), 143–166, (2003).

[11] Naoto Iwahashi, 'Robots that learn language: Developmental approach to human-machine conversations', in *International Workshop on the Emergence and Evolution of Linguistic Communication, EELC*, (2006).

[12] Simon Keizer, Mary Ellen Foster, Zhuoran Wang, and Oliver Lemon, 'Machine learning for social multiparty human-robot interaction', *TiiS*, **4**(3), 14, (2014).

[13] V Klingspor, Y Demiris, and M Kaiser, 'Human robot communication and machine learning', *Applied A.I.*, **11**, 719–746, (1997).

[14] Jens Kober, J. Andrew Bagnell, and Jan Peters, 'Reinforcement learning in robotics: A survey', *I. J. Robotic Res.*, **32**(11), 1238–1274, (2013).

[15] Evan Krause, Michael Zillich, Thomas Williams, and Matthias Scheutz, 'Learning to recognize novel objects in one shot through human-robot interactions in natural language dialogues', in *AAAI*, (2014).

[16] Cynthia Matuszek, Liefeng Bo, Luke Zettlemoyer, and Dieter Fox, 'Learning from unscripted deictic gesture and language for human-robot interactions', in *AAAI*, (2014).

[17] Michael Montemerlo, Joelle Pineau, Nicholas Roy, Sebastian Thrun, and Vandi Verma, 'Experiences with a mobile robotic guide for the elderly', in *AAAI*, (2002).

[18] Raymond J. Mooney, 'Learning to connect language and perception', in *AAAI*, (2008).

[19] Joelle Pineau and Amin Atrash, 'Smartwheeler: A robotic wheelchair test-bed for investigating new models of human-robot interaction', in *AAAI Spring Symposium on Socially Assistive Robotics*, (2007).

[20] Nicholas Roy, Joelle Pineau, and Sebastian Thrun, 'Spoken dialogue management using probabilistic reasoning', in *ACL*, (2000).

[21] Max Schwarz, Jörg Stückler, David Droeschel, Kathrin Gräve, Dirk Holz, Michael Schreiber, and Sven Behnke, 'Nimbro@home 2014 team description', in *RoboCup @ Home League*, (2014).

[22] Luc Steels, 'Social learning and verbal communication with humanoid robots', in *Humanoids*, (2001).

[23] Pawel Swietojanski, Arnab Ghoshal, and Steve Renals, 'Convolutional neural networks for distant speech recognition', *IEEE Signal Process. Lett.*, **21**(9), 1120–1124, (2014).

[24] Stefanie Tellex, Ross Knepper, Adrian Li, Daniela Rus, and Nicholas Roy, 'Asking for help using inverse semantics', in *RSS*, (July 2014).

[25] Stefanie Tellex, Thomas Kollar, Steven Dickerson, Matthew R. Walter, Ashis Gopal Banerjee, Seth J. Teller, and Nicholas Roy, 'Understanding natural language commands for robotic navigation and mobile manipulation', in *AAAI*, (2011).

[26] Jean-Marc Valin, François Michaud, Jean Rouat, and Dominic Létourneau, 'Robust sound source localization using a microphone array on a mobile robot', in *IROS*, (2003).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Embodiment, emotion, and chess: A system description

Christian Becker-Asano and **Nicolas Riesterer** and **Julien Hué** and **Bernhard Nebel**[1]

Abstract. We present a hybrid agent that combines robotic parts with 3D computer graphics to make playing chess against the computer more enjoyable. We built this multimodal autonomous robotic chess opponent under the assumption that the more life-like and physically present an agent is the more personal and potentially more effective the interaction will be. To maximize the life-likeness of the agent, a photo-realistic animation of a virtual agent's face is used to let the agent provide verbal and emotional feedback. For the latter an emotion simulation software module has been integrated to drive the agent's emotional facial expressions in parallel to its verbal utterances.

1 Introduction

Chess has been called the "Drosophila of artificial intelligence" [1] meaning that in the same way as the drosophila melanogaster has become the model organism for biological research, chess served at least for many years as a standard problem for artificial intelligence research. When in 1997 Garry Kasparov, who was ranked first at that time, lost against IBM's supercomputer "Deep Blue" [10], this problem was assumed to be solved and chess engines would nowadays outclass the best players. Altogether this triggered researchers to shift their attention to other games, such as Go. Today, for a casual chess player it can be rather frustrating to play against the computer, because he or she will lose most of the times and the computer moves its pieces with seemingly no hesitation.

Recently it was found, however, that different embodiments of the computer opponent change a human chess player's motivation to engage in a game of computer chess. These attitude changes are rooted in the humans' tendency to treat machines as social actors and this effect seems to be stronger the more human-like the machine is designed to appear [16]. With our development of the hybrid chess-playing agent MARCO, the Multimodal Autonomous Robotic Chess Opponent, we aim to investigate this research question.

The remainder of the paper is structured as follows. After discussing related work in the next section, our general motivation is explained and two research questions are introduced. Then the Elo rating will be explained together with how the employed chess engine evaluates board positions. Subsequently, MARCO's hardware components are detailed, before the interconnection of its software components is laid out. Then, the complete system is explained. Finally, we present our ideas concerning experimental protocols for evaluating MARCO. We conclude this paper with a general discussion.

[1] Artificial Intelligence Lab, University of Freiburg, 79110 Freiburg, Germany, email: [basano,riestern,hue,nebel]@cs.uni-freiburg.de

2 Related work

This section describes research projects involving chess playing robots [15, 18, 13]. They aim to answer different research questions and, therefore, they employ systems of different size and complexity.

"Gambit" is a good example for an engineer's solution to an autonomous chess-playing robotic system [15]. With their "robot manipulator system" the authors created a "moderate in cost" (i.e. 18K USD) manipulator that is able to play chess with arbitrary chess sets on a variety of boards without the need to model the pieces. Although their system does not have any anthropomorphic features, it includes a "natural spoken language interface" to communicate with the human opponent. Most importantly, "Gambit" tracks both the board and the human opponent in real time so that the board does not need to be fixed in front of the robot. With its available six degrees of freedom (DoF) and the USB camera mounted on top of its gripper the robot arm reliably grasps a wide array of different chess pieces, even if they are placed poorly. In result, it outperformed all robotic opponents at the 2010 AAAI Small Scale Manipulation Challenge. Unfortunately, no data on human players' enjoyment is available.

In contrast to the remarkable technical achievements behind the development of "Gambit", the "iCat" from Philips was combined with a DGT chess board to investigate the influence of embodiment on player enjoyment in robotic chess [13]. The authors conducted a small-scale empirical trial with the emotional iCat opponent either presented in its virtual or robotic form. Using a modified version of the GameFlow model [20], it was found that overall the virtual version is less enjoyable than the robotic one. A subsequent long term study [14] with the robotic iCat playing chess repeatedly against five children showed, however, that these children lost interest in the robot. Presumably, iCat's complete lack of any manipulation capability together with its cartoon-like appearance let the children ignore the robot completely after the initial curiosity is satisfied.

Similar to our approach, Sajó et al. [18] present a "hybrid system" called "Turk-2" that consists of a "mechanically simple" robot arm to the right of the human player and a rather simple 2D talking head presented on a computer display. "Turk-2" can analyze three emotional facial expressions, namely *sad*, *neutral*, and *happy*, and additional image processing enables the system to monitor the chess board. Interestingly, the authors decided to artificially prolong the system's "thinking time", details of which are unfortunately not reported. The transitions between the talking head's facial expressions *neutral*, *sad*, *happy*, and *bored* are controlled by a state machine that takes the human's emotion (as derived from its facial expression) and the game state into account. Similar to our approach, the talking head will change into a bored expression after some time without input has passed. An empirical study on the effect of the presence of the talking head revealed that without the talking head the players mostly ignored the robotic arm to the right of them, even when it was moving. With the talking head in front of them, however, the players not only looked at the talking head but also started smiling and laughing.

Regarding the effects of a virtual agent's facial expression of emotions on human performance in a cognitive task, an empirical trial resulted in no significant differences [8]. In addition, the study showed that for such a serious task it made no difference, if the agent's emotions were generated based on a set of hard-coded rules or by making use of a sophisticated and complex emotion simulation architecture. The authors speculate that a less cognitively demanding and more playful task might be better suited to search for such effects.

A prototype of the MARCO system has been demonstrated recently at an international conference [17] and, although conference attendees clearly enjoyed playing and loosing against the agent, several opportunities to improve the system were mentioned. The most noticeable deficiency seemed to be the use of a much too small display for presenting the agent's virtual face. Accordingly, our system now employs a much bigger display.

3 Motivation and research questions

These previous results in combination motivated us to include the following features in MARCO, our Multimodal, Autonomous, Robotic Chess Opponent:

1. A low-cost robotic arm that enables MARCO to autonomously move the chess pieces instead of having to rely on the human opponent's assistance (as in [13])
2. A custom built, robotic display presenting a highly anthropomorphic virtual agent's head to realize a hybrid embodiment combining the best of both worlds, cp. [13, 18]
3. A flexible software architecture that relies on an established emotion simulation architecture as one of its core modules (following up on [8])

The resulting MARCO system will help answering research questions that are motivated by the previous work presented above:

1. Is it more enjoyable to play chess against the robotic arm with or without the virtual agent?
2. Is it more enjoyable to play against the hybrid agent (i.e. the robotic arm with the virtual agent) when the agent expresses emotions as compared to when it remains equally active but emotionally neutral?
3. Is the most human-like and emotional agent evaluated as more social/mindful than the less complex/human-like versions of it? Does this subjective evaluation depend on how experience the human chess player is?

The first question will provide a baseline for the hardware components of our system and will be compared with those reported in [18] with regard to "Turk-2". It is not taken for granted that a more complex system will always be preferable to a simpler system from the perspective of a human player. The second question, however, is targeting the role that artificial emotions might or might not play and it is motivated by previous results [8]. Finally, MARCO allows us to tackle systematically the general question of how and when "mindfullness" is ascribed to machines [16].

4 Background and Preliminaries

4.1 Elo rating

The skill of chess players is usually measured in terms of a single integer value, the so-called Elo Rating [12]. It represents the relative strength of a player, the higher the better, and it increases or decreases with his or her chess match results. Currently, ELO rating in chess goes from 1000 (complete beginner) to 2880 (Magnus Carlsen World Champion).

Differences in the evaluations of our system might correlate with or even depend on the ELO ratings of the human players. In addition, our system might be used as a virtual coach for novice players to improve their chess skills and the ELO rating provides a standard means to compare player strength before and after training.

4.2 Chess Engine

Computer chess engines evaluate the board position using an alphabeta algorithm with a depth d given as parameter based on a number of criteria like: pieces left on the board, activity of these pieces, security of the king, etc. The greater the depth the more precise is the evaluation. The position evaluation function results in a real number e ranging from $[-\infty, +\infty]$ where 0 means that the position is equal, $-\infty$ that black is winning and $+\infty$ that white is winning. A $+1$ valuation roughly represents the advantage equivalent to a pawn, $+3$ to a knight or a bishop, and so on according to the standard valuation of chess pieces.

We denote by $e_{t,d}$ the evaluation given by the chess engine at move t with depth d. We write e when it is clear from the context. In practice, once $|e| \geq 5$ the game is more or less decided.

Our first prototype [17] was based on the TSCP chess engine [2] for its simplicity and in order to make our results comparable to previous work on the iCat playing chess [13], for which the same engine was used. The communication between the user and the TSCP chess engine is handled by the XBoard Chess Engine Communication Protocol [3]. Originally implemented as a means to facilitate communication between the GNU XBoard Chess GUI and underlying chess engines, this plain text protocol allows for easy information exchange in a human readable form.

Our modular software architecture allows us, however, to plug in other chess engines. The more advanced Stockfish chess engine [4], for example, would allow us to adjust the strength of MARCO's play dynamically.

5 Hardware components

The complete setup is presented in Figure 1. The hardware used comprises a custom designed, 15.6 inch pan-tilt-roll display to present the virtual agent's face, a robotic arm to the right of the agent to move the chess pieces, and a digital chess board (DGT USB Rosewood) with a chess clock. Each of these components will be described next.

5.1 The pan-tilt-roll agent display

The pan-tilt-roll display component features a 15.6 inch upright TFT LCD display with a physical resolution of 1920×1080 pixels and 18bit color depth, cp. Fig. 1. It is positioned opposite of the human player to give the impression of the virtual agent overlooking the complete chess board. Three Dynamixel AX-12A servos (cp. Fig. 2(a)) are connected to USB2Dynamixel interface to allow for control over the display's orientation during the game along all three axes. Thereby, for example, the agent can follow its own arm's movements dynamically as presented in Fig. 2(b).

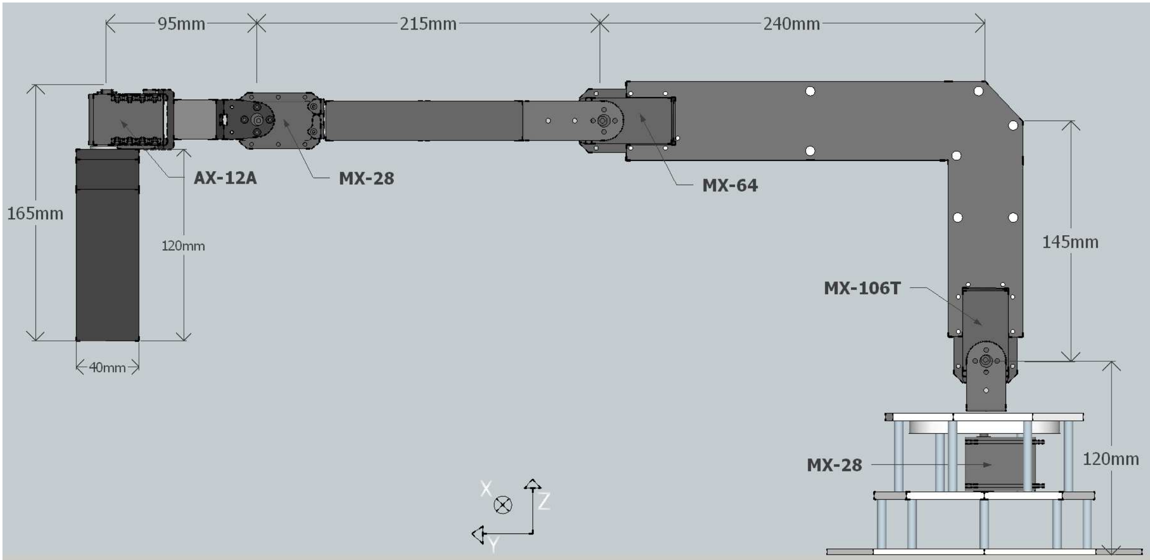


Figure 3. A schematic of the robotic arm with annotations of link lengths and Dynamixel servos used for each joint position

Figure 1. The pan-tilt-roll agent display, the robotic arm, and the digital chess board together with the digital chess clock

5.2 The robotic arm

The hybrid agent's robotic arm is a modification of the "WidowX Robotic Arm Kit Mark II" [5] available from Trossen Robotics. Apart from the rotational base all other parts needed to be extended to allow the agent to pick-and-place all pieces on any of the 64 squares of the board. The upper arm was extended to measure $240mm$, the forearm to measure $215mm$ and the gripper needed to be prolonged to $120mm$ (cp. Fig. 3). These extensions for the arm as well as the extra parts to realize the display mount were printed with a MakerBot 3D printer. Five Dynamixel servos move the robot's arm, cp. Fig. 3. For the base and wrist two MX-28 servos are used. An MX-64 servo moves the robot's elbow and an MX-106 servo its shoulder. The modified gripper is opened and closed by an AX-12A servo, cp. Fig. 4. It can reliably pick-and-place all Staunton chess pieces on the DGT board regardless of their height or size.

(a) Detail view (b) Front view

Figure 2. Pan-tilt-roll mount of the 15.6 inch display presenting the virtual agent's face

5.3 The DGT digital chess board

The DGT chess board is a wooden board with standard Staunton pieces and $55mm \times 55mm$ squares. Each piece is equipped with a unique RFID chip that makes it recognizable. The board is con-

(a) Open state (b) Closed state

Figure 4. The two states of the robot's custom designed gripper picking up a white bishop

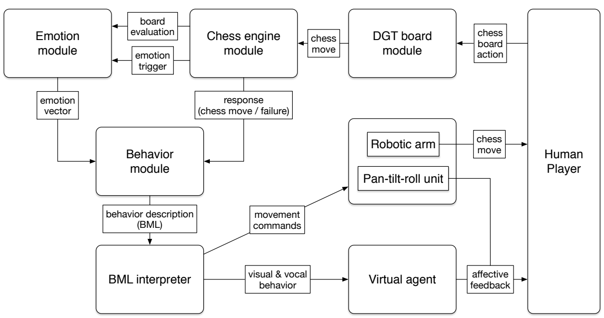


Figure 5. An outline of the software modules and their connections

nected to the computer with a USB cable, and it transmits the position in FEN format to the DGT board module every time a change is performed.

6 Software components

Except for the external MARC framework (see Section 6.3), all components are implemented in C++ using Qt5 [7] in combination with the Robot Operating System (ROS; [6]) to achieve a modular design and cross-platform functionality. The hardware components (i.e. the DGT chess board and the Dynamixel servos) are encapsulated into ROS nodes to establish a flexible communication infrastructure.

6.1 Overview of system components

The following five main software components can be distinguished, which are connected by the ROS message protocol (cp. Fig. 5):

- A DGT board module to detect moving pieces on the physical chess board
- A Chess engine model for position evaluation and chess move calculation
- An Emotion module to simulate MARCO's emotions
- A Behavior module to integrate the chess move with emotional states into a behavior description
- A Behavior Markup Language (BML) Interpreter to prepare the multimodal realization of the behavior
- Robotic components to move the chess pieces on the board and control the virtual agent's pan-tilt-roll unit
- The MARC framework to create the agent's visual appearance on the display

When the human player (cp. Fig. 5, right) performs her move, the DGT board module recognizes the change on the board, derives the move information by comparing the current board configuration with the previous one, and sends this information to the chess engine module. Here, the chess move is verified for correctness and either (1) a failure state, or (2) the chess engine's move is transmitted as MARCO's response to the behavior model. The board evaluation function of the chess engine also provides the emotion module with input. After the emotion module integrated the board evaluation into the agent's emotion dynamics (see Section 6.2), it concurrently updates the behavior module with a vector of emotion intensities. The behavior module integrates the emotional state information with the move calculation into a behavior description in BML [21]. This description is then interpreted by the BML interpreter to drive the virtual agent's visual and vocal behavior as well as the robotic component's actions. While the robotic arm starts to execute the agent's chess move, the pan-tilt-roll unit moves the display to realize affective feedback in combination with the virtual agent's facial expressions.

6.2 Deriving emotional states

The emotion module (cp. Fig. 5) comprises the WASABI Affect Simulation Architecture [9] to simulate the agent's dynamically changing emotional states. As input WASABI needs *valenced impulses* and expectation-based emotions (e.g., *surprised* and *hope*) need to be *triggered* before they can gain positive intensity.

6.2.1 Emotion dynamics

WASABI is based on the idea that emotion and mood are tightly coupled. The term "mood" refers to a relatively stable background state, which is influenced by emotion arousing events, but changes much more slowly as compared to any emotional state. An "emotion", in contrast, is a short-lived psychological phenomenon that more directly impacts behavior than a mood does.

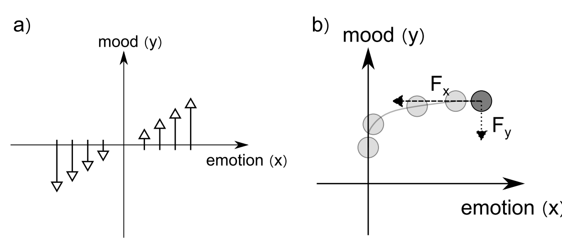


Figure 6. The emotion dynamics of WASABI with (a) the influence of emotional valence on mood, and (b) the effect of the two independent mass-spring systems on the development of the agent's emotional state over time (indicated by the half-transparent circles)

Taking these differences and commonalities as cue, WASABI simulates the positive and negative effects that emotional valence has on mood, cf. Fig. 6a. In addition, mood and emotion are driven back to zero by two forces independently exerted from two mass-spring systems. Notably, the respective spring constants are set such that the resultant force F_x is always greater that the resultant force F_y, because emotions are longer lasting than mood, cp. Fig. 6b.

MARCO's emotional state as represented in Fig. 6b by the circles is updated with 50Hz letting it move through the space over time. The x and y values are incessantly mapped into PAD space to allow for categorization in terms of emotion labels (cp. Fig. 7; see also [9]).

This dynamic process is started by the arrival of a *valenced impulse* from outside of WASABI that instantaneously changes the emotion value (x) either in the positive or negative direction. How these impulses are derived from the progression of the game is described next.

6.2.2 *Valenced impulses*

The chess engine module continuously calculates board evaluations e_t (at times t during the game). These are converted into *valenced impulses* $val(e_t)$ according to Equation 1.

$$val(e_t) = k \times tanh\left(\frac{e_t}{r}\right) \tag{1}$$

Here, k is a scaling factor and by increasing the denominator $r \in [1, \infty]$ the skewness of the hyperbolic tangent is reduced until a quasi-linear mapping ($val(e_t) = k \times e_t$) is achieved. The hyperbolic tangent is introduced to let us emphasize small values of e_t relative to bigger values of e_t.

For example, choosing $k = 50$ and $r = 2$:

$$\begin{gathered} val(e_t) = 50 \times tanh\left(\frac{e_t}{2}\right) \in (2.5, 25], \\ \forall e_t \in \{x \in \mathbb{R} \mid 0.1 \leq x < 1.1\} \end{gathered} \tag{2}$$

Thus, with these constants any value of e_t between 0.1 and 1.1 results in a weak to medium valenced impulse. Observe that $|val(e_t)| \cong 50, \forall e_t \in \{x \in \mathbb{R} \mid |x| > 5\}$, meaning that a winning (or loosing) board configuration results in the maximum impulse of 50 (or minimum impulse of -50, respectively).

Depending on who plays white, the sign of the scaling factor k is adjusted as to map favorable board positions for MARCO to positively valenced impulses and vice versa. That is, if MARCO plays white k is positive, otherwise it is negative. For the time being, MARCO always plays white letting it perform the first half-move.

Inside the emotion module the *valenced impulses* drive the concurrent simulation of the agent's emotion dynamics. In summary, a positive (negative) impulse has the short term effect of increasing (decreasing) the agent's *emotional valence*, which in turn influences the agent's *mood* in the same direction as a long term effect. A simple mathematical transformation into *pleasure* ($P = \frac{x+y}{2}$) and *arousal* ($A = |x|$) values is performed and the emotion module then uses the PAD space (cf. Fig.7) to categorize the agent's emotional state in terms discrete emotions and their intensities. The *dominance* value is changed in accordance with whether it is MARCO's turn ($D = 1$) or not ($D = 0$). Finally, the resulting set of emotions with positive intensities are transmitted to the behavior module.

6.2.3 *Mapping onto discrete emotions*

In its default configuration, WASABI simulates the primary emotions *annoyed*, *angry*, *bored*, *concentrated*, *depressed*, *fearful*, *happy*, *sad*, and *surprised* as well as the secondary emotions *relief*, *fears-confirmed*, and *hope*; cp. Fig. 7. Five of these 12 emotions (*fearful*, *surprised*, *relief*, *fears-confirmed*, and *hope*) rely on an agent's ability to build expectations about future events, i.e., they are so-called *prospect-based emotions*. For example, one is only surprised about an event, if it is contrary to one's previous expectations, or one fears future events, only if one has reason to expect that bad event is about to happen [9]. Accordingly, in WASABI each of these emotions is configured with zero *base intensity* and needs to be *triggered* (cp. "emotion trigger" in Fig. 5) to give them a chance to gain positive intensity.

With respect to chess, our system evaluates the available moves for its opponent. MARCO is able to realize, whenever its last move was less good than previously evaluated, because at time t the evaluation reaches one level deeper into the search tree than at time $t-1$. Accordingly, MARCO might start to fear that the human opponent realizes her opportunity as well. If the evaluation of the situation after the opponent's move is stable, then MARCO's fears are confirmed: the opponent made the right move. On the other hand, if the evaluation comes back to what it was before, i.e., before MARCO made its last move, then the opponent missed the opportunity and MARCO is relieved. The evaluation can be in between these two values and in that case, the agent is neither relieved nor sees its fears confirmed. Nevertheless, the emotion module still receives the negative *valenced impulse* derived from the drop. Formally, Table 1 provides details on how the changing evaluations trigger prospect-based emotions in WASABI.

Notably, the value e_t represents the future directed evaluation of the situation from the robot's perspective. For example, the formula $e_{t-1} - e_t > \epsilon$ lets the behavior trigger *fear* whenever a significant drop in the evaluation function appeared from the previous move to

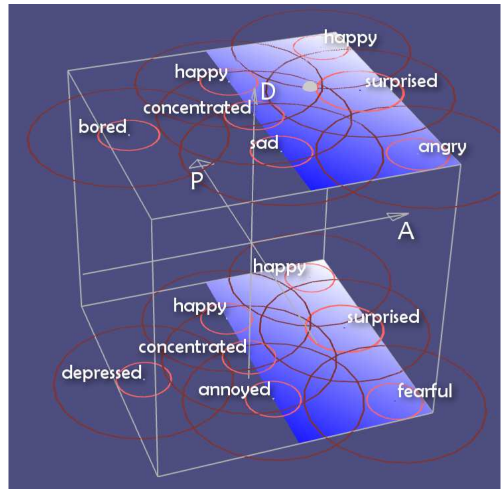


Figure 7. The PAD-space of primary and secondary emotions in WASABI. The primary emotions are distributed as to cover all areas of PAD space. For each of them an activation threshold (outer ring) and a saturation threshold (inner ring) is defined. The two shaded areas represent the distribution of the secondary emotion *hope* in the dominant and submissive subspace, after it was triggered. The grey half-sphere represents MARCO's dynamically changing emotional state. Thus, in this example MARCO would be mildly *happy*, a bit *concentrated*, and quite *hopeful*. If surprise were triggered as well in this moment, MARCO would also be *surprised* to a certain extend.

trigger	*if..*
fear	$e_{t-1} - e_t > \epsilon$
surprise	$\lvert e_{t-1} - e_t \rvert > \epsilon$
fears-confirmed	$fear_{t-1} \wedge (e_{t-1} - e_t < \epsilon)$
hope	$e_{t,d} - e_{t,d-2} > \epsilon$
relief	$fear_{t-1} \wedge (e_t - e_{t-2} < \epsilon)$

Table 1. The conditions under which the prospect-based emotions are triggered in WASABI based on the changes of evaluations over time with ϵ and depth d as custom parameters

Figure 8. The virtual agent expressing *anger*, *neutral*, and *joy* (left to right)

the current one. That is, MARCO realizes at time t that the future seems much worse than evaluated before (in time $t-1$). If subsequently, after the next half-move in $t+1$, the value e_{t-1} turns out to have been correct in the light of the new value e_t (or the situation got even worse than expected), then *fears-confirmed* will be triggered. On the contrary, if it turned out to be much better than expected, *relief* will be triggered. *Surprise* is always triggered when the evaluation changes significantly from one half-move to the next. Finally, *hope* is triggered whenever not taking the full depth of the search tree into account would mean that the key move in the position is hard to reach (requires a computation at depth at least *d*).[2]

6.2.4 *The emotion vector as input for the behavior module*

It is important to note that, in addition to an emotion being triggered, the *pleasure*, *arousal*, and *dominance* (PAD) values driven by the emotion dynamics must be close enough to that emotion for it to become a member of the *emotion vector* with positive intensity, cp. Fig. 7. Thus, although *surprise* will always be triggered together with *fear*, they will not always both be present in the *emotion vector*, because they occupy different regions in PAD space.

From the *emotion vector* the emotion with the highest intensity is compiled into the BML description driving the MARC framework. The agent comments on particular events like, for example, complimenting the player after it lost a game or stating that the position is now simplified after exchanging the queen.

6.3 The virtual agent provided by the MARC framework

The MARC framework [11] is used to animate the virtual agent, which is presented on the 15.6 inch pan-tilt-roll display facing the human player. The emotional facial expressions (see Fig. 8 for examples) that are provided as part of the BML description are combined inside the MARC framework to create lip-sync animations of emotional verbal utterances. Thanks to the integration of the open-source text-to-speech synthesis OpenMARY [19] the agent's emotion also influences the agent's auditory speech.

7 Conclusions and future work

This paper detailed the software and hardware components behind MARCO, a chess playing hybrid agent equipped with a robotic arm and a screen displaying a virtual agent capable of emotional facial expressions. A first prototype of the system was demonstrated at an international conference [17] and the experiences gained let to improvements both on concerning the hard- and software components.

Although a limited set of concrete agent behaviors has proven to be fun for the conference participants, we still need to design many more of them. For example, we need to decide which kind of comments are to be given with which timing during the game and how virtual gaze and robotic head movements are to be combined to give the impression of a believable, hybrid agent.

In order to answer the initially stated two research questions, we plan to conduct a series of empirical studies. At first, one group of participants will play against MARCO with the pan-tilt display turned off. Nevertheless, the invisible agent's comments will remain audible in this condition. In the second condition, another group of participants will play against MARCO with an unemotional agent presented on the robotic display. For the third condition, a group of participants will play against the WASABI-driven agent. In all three conditions, player enjoyment will be assessed using the GameFlow [20] questionnaire and video recordings of the human players will be analyzed inspired by [18]. We expect to find significant differences between conditions with the most complete setup (condition three) being most fun for the players.

Nass and Moon claim that imperfect technologies mimicking human characteristics might even increase "the saliency of the computer's 'nonhumanness'." [16, p. 97] In line with their ideas and in addition to the approach outlined above, we plan to compare human-human interaction with human-agent interaction when competing in chess to measure and incessantly improve MARCO's level of human-likeness. This will help to understand how human behavior might be split into computationally tractable components and then realized in robotic agents to improve human-computer interaction.

REFERENCES

[1] http://aitopics.org/topic/chess.
[2] http://www.tckerrigan.com/Chess/TSCP.
[3] http://www.gnu.org/software/xboard/engine-intf.html.
[4] https://stockfishchess.org/.
[5] http://www.trossenrobotics.com/widowxrobotarm.
[6] http://www.ros.org/.
[7] Qt: Cross-platform application and UI framework. http://qt-project.org/, May 2014.
[8] C. Becker-Asano, P. Stahl, M. Ragni, J.-C. Martin, M. Courgeon, and B. Nebel. An affective virtual agent providing embodied feedback in the paired associate task: system design and evaluation. In *Proc. of the 13th. Intl. Conf. on Intelligent Virtual Agents (IVA 2013)*, pages 406–415, Edinburgh, UK, August 2013.
[9] C. Becker-Asano and I. Wachsmuth. Affective computing with primary and secondary emotions in a virtual human. *Autonomous Agents and Multi-Agent Systems*, 20(1):32–49, January 2010.
[10] M. Campbell, A. H. Jr., and F. hsiung Hsu. Deep blue. *Artificial Intelligence*, 134(1 – 2):57 – 83, 2002.

[2] An evaluation function is usually set up to an even number, thus the last level of the search tree equals the last two half-moves.

[11] M. Courgeon, J.-C. Martin, and C. Jacquemin. MARC: a Multimodal Affective and Reactive Character. In *Proc. 1st Workshop on AFFective Interaction in Natural Environements*, 2008.

[12] A. E. Elo. *The rating of chessplayers, past and present*. Arco Pub., New York, 1978.

[13] I. Leite, C. Martinho, A. Pereira, and A. Paiva. icat: an affective game buddy based on anticipatory mechanisms. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems - Volume 3*, AAMAS '08, pages 1229–1232, Richland, SC, 2008. International Foundation for Autonomous Agents and Multiagent Systems.

[14] I. Leite, C. Martinho, A. Pereira, and A. Paiva. As time goes by: Long-term evaluation of social presence in robotic companions. In *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, pages 669–674, Sept 2009.

[15] C. Matuszek, B. Mayton, R. Aimi, M. Deisenroth, L. Bo, R. Chu, M. Kung, L. LeGrand, J. Smith, and D. Fox. Gambit: An autonomous chess-playing robotic system. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 4291–4297, May 2011.

[16] C. Nass and Y. Moon. Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, 56(1):81–103, 2000.

[17] N. Riesterer, C. Becker Asano, J. Hué, C. Dornhege, and B. Nebel. The hybrid agent MARCO. In *Proc. of the 16th Intl. Conf. on Multimodal Interaction*, pages 80–81. ACM, 2014.

[18] L. Sajó, Z. Ruttkay, and A. Fazekas. Turk-2, a multi-modal chess player. *International Journal of Human-Computer Studies*, 69(7–8):483–495, 2011.

[19] M. Schröder. OpenMARY sources. https://github.com/marytts/marytts, April 2013.

[20] P. Sweetser and P. Wyeth. Gameflow: a model for evaluating player enjoyment in games. *Computers in Entertainment (CIE)*, 3(3):3–3, 2005.

[21] H. Vilhjálmsson, N. Cantelmo, J. Cassell, N. E. Chafai, M. Kipp, S. Kopp, M. Mancini, S. Marsella, A. N. Marshall, C. Pelachaud, Z. Ruttkay, K. Thóisson, H. Welbergen, and R. J. Werf. The behavior markup language: Recent developments and challenges. In *Intelligent Virtual Agents*, volume 4722 of *LNCS*, pages 99–111. Springer Berlin Heidelberg, 2007.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Towards a Child-Robot Symbiotic Co-Development: a Theoretical Approach

Vicky Charisi[1], Daniel Davison[1], Frances Wijnen[2], Jan van der Meij[2], Dennis Reidsma[1], Tony Prescott[3], Wouter van Joolingen[4], Vanessa Evers[1]

Abstract. One of the main characteristics for an effective learning is the possibility for learners to choose their own ways and pace of learning, according to their personal previous experiences and needs. Social interaction during the learning process has a crucial role to the skills that learners may develop. In this paper, we present a theoretical approach, which considers relevant theories of child's development in order to proceed from a child-child collaborative learning approach to a child-robot symbiotic co-development. In this symbiotic interaction, the robot is able to interact with the learner and adapt its behaviours according to the child's behaviour and development. This sets some theoretical foundations for an on-going research project that develops technologies for a social robot that facilitates learning through symbiotic interaction.

1 INTRODUCTION

This paper discusses the conceptualization and some initial investigation of children's collaborative learning through symbiotic child-robot interaction in a specific educational setting. According to Douglas [1], biologist Heinrich Anton de Bary used the term "symbiosis" in 1879 to describe any association between *different* species. In this context, symbiotic learning describes the process, during which members of a team mutually influence each other resulting in an alteration of their behaviour. However, relationships among members may sustain imbalances. In order to support symbiotic interactions in learning, special considerations should be given to the orchestration of the relationships and the process between members of the team, from which they all benefit. The core motivating principle of symbiosis and the collaboration within it is reciprocity. Thus, learning emerges through a harmonized openness, responsiveness and adaptation. Elements of this kind of interaction may appear also in collaborative learning settings, which may not be especially designed for symbiotic interactions. Identifying elements of symbiotic interaction in children's collaborative learning may provide us with features for a more effective interaction design and for the design of robot behaviours as the child's co-learner.

In the following sections, we describe some constructivist aspects of child learning focusing on the need for learners to take responsibility for the regulation of the form and pace of learning. We then describe how symbiotic interaction can provide a theoretical and practical framework for understanding child-robot inter-dependence.

[1] Dept. of Electrical Engineering, Mathematics and Computer Science, HMI, University of Twente, Enschede, the Netherlands.
Email:{v.charisi, d.p.davison, d.reidsma, v.evers}@utwente.nl

[2] Faculty of Behavioural Management and Social Sciences, ELAN, University of Twente, Enschede, the Netherlands, Email: {f.m.wijnen, j.vandermeij}@utwente.nl

[3] Sheffield Robotics, University of Sheffield, Mappin Street, Sheffield, UK. Email: t.j.prescott@sheffield.ac.uk

[4] Freudenthal Institute for Science and Mathematics Education, Faculty of Science, Utrecht University, Utrecht, the Netherlands, Email: {w.r.vanjoolingen@uu.nl}

2 ASPECTS OF CHILDREN'S LEARNING PROCESSES

According to Foston and Perry [2], learning is a constructive activity that occurs through the interaction of individuals with their surroundings. Stages of development are understood as constructions of the active re-organization of learner's knowledge. This view builds on the constructivist framework of Piagetian developmental theory [3] according to which learning is a dynamic process comprising successive stages of adaption to reality, and during which learners actively construct knowledge by creating and testing their own theories and beliefs.

Two aspects of Piaget's theory underpin the pedagogical approach adopted here: First, an account of the four main stages of cognitive development through which children pass [4]. Since their birth, children go through (i) the sensori-motor stage (0-2 years), (ii) the pre-operational (2-7 years), (iii) the concrete operational (7-12 years) and (iv) the formal operational stage (12 years and onwards). For this project, we consider children in the age group between 7 and 12 years. During this stage, children are able to imagine "what if" scenarios, which involve the transformation of mental representation of things they have experienced in the world. These operations are "concrete" because they are based on situations that children have observed in the environment.

Second, an account of the mechanisms by which cognitive development takes place [5], which we consider in relation to environmental, social and emotional elements of child's development. These mechanisms describe how children actively construct knowledge by applying their current understanding.

2.1 Learning as a dynamic process

According to Piaget's classic constructivist view, learning occurs in a sequence of stages from one uniform way of thinking to another. Cognitive conflict, arising from discrepancies between internal representations and perceived events, functions as the motivating force for changing from concrete modes of thinking to more abstract forms. Although these stages relate to the child's genetic predispositions and biological development, environmental factors affect the transition from one stage to the next in complex ways. However, since Piaget first defined his framework it has been recognized that developmental transitions are not necessarily age specific events, but it occurs within an age range that can differ from child to child [6]. Additionally, the relationship between child development and the context in which this occurs, is bi-directional which results in a dynamical, iterative process; children affect and, simultaneously, they are affected by factors of their environment [7]. This can happen either in informal settings [8] which support tinkering and learning by doing or by following more formal and standardized processes, such as the inquiry cycle process [9], which will be described in 2.1.2 of this paper.

2.1.1 Child and the natural need for learning through exploration

In order for a child to be strongly engaged with a task it has to be meaningful for them. Since children have an inherent motivation to explore and understand their surroundings, the relevance of the task will stimulate their curiosity and willingness for exploration. Science education provides a formal learning setting that should share some of the characteristics of informal settings in order to help children acquire new concepts and develop transferable skills. Building on constructivist principles, children's natural enthusiasm for play can be a key factor in learning. During play, children can explore the real world, logically organize their thoughts, and perform logical operations [10]. However, this occurs mainly in relation to concrete objects rather than abstract ideas [8]. Children are also able to reflect on their intentional actions which may result in a self-regulated process of change [11].

2.1.2 Inquiry cycle: a systematic process of learning

'Inquiry is an approach to learning that involves a process of exploring the natural or material world, and leads to asking questions, making discoveries, and rigorously testing those discoveries in the search for new understanding. Inquiry should mirror as closely as possible the enterprise of doing real science' [12] (p.2). The main claim of inquiry learning, in relation to science learning, is that it should engage learners in scientific processes to help them build a personal scientific knowledge base. They can then use this knowledge to predict and explain what they observe in the world around them [13]. Thus, having as a starting point child's tendency for informal exploration, with developmental appropriate scaffolding, children develop their scientific thinking. This transferable skill can then facilitate child learning in different contexts.

There are many models that represent the processes of inquiry, but all include the processes of (1) *hypothesis generation* in which learners formulate their ideas about the phenomena they are investigating, (2) *experimentation* in which children perform experiments to find evidence for rejection or confirmation of their hypotheses and (3) *evidence evaluation* in which learners try to find logical patterns in their collected data and to interpret this data to form a conclusion [14, 15].

Banchi and Bell [9] describe a four-level continuum to classify the levels of inquiry in an activity, focusing on the amount of information and guidance that is presented to the learner [9, 16]:

Confirmation inquiry: In this form of inquiry learners are provided with the research question, method of experimentation and the results that they should find. This is useful if, for example, the goal is to introduce learners to the experience of conducting investigations or to have learners practice a specific inquiry skill such as collecting data.

Structured inquiry: Here, the question and procedure are still provided but the results are not. Learners have to generate an explanation supported by the evidence they have collected. In this case learners do know which relationship they are investigating.

Guided inquiry: In this form learners are provided only with the research question. Learners need to design the procedure to test their question and to find resulting explanations.

Open inquiry: This is the highest level of inquiry. Here, learners have the opportunities to act like scientists, deriving questions, designing and performing experiments, and communicating their results. This level requires the most scientific reasoning and is the most cognitive demanding. This low- to higher-level continuum of inquiry is important to help learners gradually develop their inquiry abilities [9]. The obtained inquiry skills are transferable to other contexts.

2.2 The zone of proximal development (ZPD)

The level of potential development is the level at which learning takes place. It comprises cognitive structures that are still in the process of maturing, but which can only mature under the guidance of or in collaboration with others. Vygotsky [17] distinguished between two developmental levels: the level of actual development and that of potential development. The actual development is the level, which the learner has already reached and she can solve problems independently. The level of potential development, which is also known as the *Zone of Proximal Development (ZPD),* describes the place where child's spontaneous concepts meet the systematic reasoning under the guidance or in collaboration with others [18]. In that way, Vygotsky argues that the interpersonal comes before the intrapersonal. This is considered to be as one of the fundamental differences between Vygotsky's conceptualization of child development and that of Piaget.

Learning takes place within the ZPD and here a transition occurs in cognitive structures that are still in the process of maturing towards the understanding of scientific concepts. The level of potential development varies from child to child and is considered a fragile period for child's social and environmental support through the educational praxis. In this context, Vygotsky introduced the notion of 'scaffolding', to describe the expansion of the child's zone of proximal development that leads to the construction of higher mental processes [19]. However, only if we define what causes the expansion of ZPD, we will be able to provide appropriate scaffolding for learners. Siegler [20], for example, has highlighted the question of what causes change in learning mechanism and he concluded that seemingly unrelated acquisition are products of the same mechanisms or mental process. Scaffolding is considered a core element for the support of child's mental changes in the context of collaborative learning.

2.3 Collaborative learning

Rogoff's [21] definition of collaboration includes mutual involvements and engagement and participation in shared endeavours, which may or may not serve to promote cognitive development. This broad definition allows for flexibility regarding its interpretation and it is adjustable into different contexts. For the present research, we use this definition as a basis for our theoretical approach for collaboration in the context of learning.

Vygotsky [17] emphasized the importance of social interaction with more knowledgeable others in the zone of proximal development and the role of culturally developed sign systems that shape the psychological tools for thinking.

In addition to the development of their cognitive skills, children's social interactions with others during the learning process may trigger their meta-cognitive skills, as well. Providing explanations during collaboration in which children reflect on the process of their learning (meta-cognitive skills) leads to deeper understanding when learning new things [22, 23]. There are two forms of explanation: (1) self- explanation, which refers to explanation of the subject of interest to oneself, and (2) interactive explanation, which refers to explanation to another person [24]. In both cases, the presence of a social partner facilitates children's verbalization of their thinking. However depending on the type of the social partner, children may exhibit different behaviours, which relate to different kind and quality of learning.

The following sections describe two different types of social partners as mediators for children's learning to occur.

2.3.1 Child – tutor

With regard to adult-child interactions, Wood et *al.* [25] defined tutoring as 'the means whereby an adult or 'expert' helps somebody who is less adult or less expert' (p.89). Receiving instructions from a tutor is a key experience in childhood learning (ibid.). This definition of tutoring implies a certain mismatch in the knowledge level between the parties involved, in such a way that the tutor has superior knowledge or skill about a subject which is then passed on to a child via tutoring mechanisms.

2.3.2 Child – child

In combination with tutoring, peer learning has been defined by Topping [26] as 'the acquisition of knowledge and skill through active helping and supporting among status equals or matched companions' (p.1). Topping continues to describe that peer learning 'involves people from similar social groupings who are not professional teachers helping each other to learn and learning themselves by so doing' [26]. This learning method has proven to be very effective amongst children and adults and has been widely researched over the past decades. Peer learning assumes a matched level of initial knowledge of both parties. In ideal peer learning situations, both parties will increase their knowledge levels at a similar pace through collaborative learning mechanisms.

2.4 Emotional engagement and social interaction (in learning)

The importance of positive feelings during the learning process has been reported as crucial [27]. They promote the individual's openness to new experiences and resilience against possible negative situations [28]. It has been reported that dynamic behaviours involve reciprocal influences between emotion and cognition [29]. For instance, emotions affect the ways in which individuals perceive the reality, pay attention and remember previous experiences as well as the skills that are required for an individual to make decisions.

3 SYMBIOTIC INTERACTION

The educational and developmental theories outlined in the previous sections describe various forms of collaborative learning. Social interaction between learners is emphasised as an important factor in successful collaborative learning, where both students co-develop at a complementary pace through shared experiences.

Within the context of this co-development we define *symbiotic interaction* as the dynamic process of working towards a common goal by responding and adapting to a partner's actions, while affording your partner to do the same.

The fundamental requirements for team collaboration have been discussed in detail by Klein and Feltovich [30]. They argue that in order to perform well on *joint activities*, or collaborative tasks, there must be some level of *common ground* between teammates. These concepts have been introduced by Clark [31] to describe the intricate coordination and synchronization processes involved in everyday conversations between humans.

Common ground between team participants is the shared mutual knowledge, beliefs and assumptions, which are established during the first meeting and continuously evolve during subsequent interactions. A strong common ground can result in more efficient communication and collaboration during joint activity, since a participant can assume with relative safety that other participants understand what she is talking about without much additional explanation [30].

Klein and Feltovich [30] argue that in order for a task to qualify for effective joint activity, there must firstly be an *intention* to cooperate towards a common goal and secondly the work must be interdependent on multiple participants. As long as these preconditions are satisfied, a joint activity requires *observable, interpretable* and *predictable* actions by all participants. Finally, participants must be open to *adapt* their behavior and actions to one another. The different processes of the joint activity are choreographed and guided by clear *signaling of intentions* between participants and by using several coordination devices such as agreement, convention, precedent and salience.

3.1 Intention to act towards a common goal

An important precondition for symbiotic interaction is the *awareness* of a certain common goal, and a clear *intention* to work towards this goal. During the process of establishing and maintaining common ground, both parties will (implicitly or explicitly) become aware of the goals of the other. Maintaining common ground relies on being able to effectively signal your intent to a partner, while at the same time interpreting and reacting to the intent of his or her actions [30].

3.2 Observability of actions and intentions

Equally important to being able to effectively *signal* intent is the ability of the partner to *observe* and *interpret* this intent. A sense of *interpredictability* can be achieved when such signals can be naturally and reliably generated, observed and interpreted by both partners. A healthy level of interpredictability between partners can contribute to an increased common ground and mutual trust between partners [30].

3.3 Interpredictability, adaptability and trust

Within the context of an interaction, predictability means that one's actions should be predictable enough for others to reasonably rely on them when considering their own actions. Over the course of an interaction, certain situations arise which allow a person to estimate the predictability of a partner's actions, or in other words, the amount of *trust* you place in the predictability of your partner. Simpson [32] argues that in human-human interaction, trust levels are often established and calibrated during trust-diagnostic situations "in which partners make decisions that go against their own personal self-interest and support the best interests of the individual or the relationship" [32]. This *willingness* to act predictably and *adapt* one's behavior to support a partner's best interests is a key component of building mutual trust and supporting a symbiotic relationship [33].

In summary, an effective joint activity relies on signaling, observing and interpreting the intent of actions towards a common goal. By establishing a strong common ground, both partners achieve a level of interpredictability. An important factor in building trust is to expose a willingness to act predictively and adapt one's behavior to match the common goals shared with a partner.

4 CHILD-ROBOT INTERACTION

The work reported in this paper is part of a project on social robots in learning scenarios. Social interaction with a robot affects the child's independence during the learning process. Robots can take either end of the spectrum depending on its role, in other words, it can be either tutor-like or peer-like for child learning [34]. Depending on the amount of support needed for the child's learning, the robot might adapt its role to fit this need, shifting either more towards the tutor or the peer role. This adaptive behavior fits the theories on symbiotic interactions outlined above. Together with clear signaling of intents, which contribute to an increased level of predictability, it is this adaptability that proves to be an important factor in building a long-term symbiotic relationship.

Belpaeme et al. [35], for example, have reported the importance of adaptive behavior of the robot when it interacts with children with diabetes. In this study, researchers adapted robot behaviour according to children personality (extroverted / introverted) and to the difficulty level of the task. They concluded that adaptation to user characteristics is an effective aid to engagement.

In the context of the learning process, a robot may adapt its behavior to the child's cognitive, social and emotional characteristics with a purpose to facilitate the expansion of children's zone of proximal development. Thus, the robot can scaffold the process of change by adapting its behaviour according to the user. It shows its awareness and willingness to be influenced by others. The robot then will adapt to the child's next level in order to contribute to the iterative process of development. In this way, we create a learning context based on symbiosis of the child and the robot.

5 FUTURE AGENDA

Inspired by the insights derived from the previously introduced theoretical framework for co-development in learning, we outline our future goals, which focus on the elaboration of aspects of this framework and explore its utility for designing robot-child interactions for inquiry learning. To conclude this paper we briefly describe a contextual analysis we are performing to validate the framework in the specific pedagogic setting of inquiry learning. Thereafter we briefly present some of our ideas for future experiments.

5.1 Some first insights from a contextual analysis

An initial contextual analysis is being performed based on observations of twenty-four children who are working in pairs on a balance beam task. The balance beam task is a specific implementation of a type of structured inquiry learning. Using the balance beam children investigate the weight of several provided objects, exploring both the influence of weight rations and the distance of the object to the pivot.

The setting for this contextual analysis was as follows: a total of 11 pairs of two children (aged 6-9 years) received a structured assignment, which they could complete by using the balance beam that was presented. This assignment was designed according to the processes of structured inquiry (e.g. hypothesis generation, experimentation, evidence evaluation). The children could place pots that differed in weight on different places on the balance, make predictions about what would happen to the balance (tip left, tip right, or stay in equilibrium), perform experiments by removing wooden blocks that held the balance in equilibrium, observe what happened with the balance and draw conclusions about the variables that influence the balance (weight, distance). These procedures were videotaped and than annotated. These annotations are not yet fully analysed, but a few first indications will be described here.

1. It appeared that children who followed the steps of the assignment correctly were engaging in the different processes that are typical for inquiry learning, and were interacting with each other about the process and the outcome of the task.
2. Most children were able to identify the influence of the two variables (weight and distance) on the balance eventually.
3. Several children asked for additional guidance from the experimenter during the task.

These first insights from the contextual analysis have been taken into account for our next steps for the design of child-robot interaction in the same context. We observed that children in this age may follow the inquiry process during the activity. However, in order for them to reflect on this process, verbalize their thoughts and explain the scientific phenomenon under investigation, they needed the support from a social partner. The teacher facilitated child's process by different types of interaction, such as supporting children's inquiry process by probing questions or asking for explanations and summarizations. In addition to the verbal interaction, we considered non-verbal cues of social interactions that appeared during this contextual analysis. The emerging types of social interactions have informed our design for future experiment on child-robot interaction.

5.2 Planned experiments

Our next steps include two experiments on child-robot interaction. In the first experiment we will focus on the influence of a social robot on explanatory behavior. Explanatory behavior includes the verbalization of scientific reasoning of the child.

The experiment is comprised of two conditions. In the experimental condition the child will be working on an inquiry assignment with the robot. The background story of the robot is that he comes from another planet. He has an assignment from his teacher to study the effects of balance on earth. The robot wants to explore this phenomenon with like-minded people: children. The robot is presented as a peer learner but he does have well-developed inquiry skills. Therefore, the robot will provide instructions and ask questions to help learners explore the phenomenon of balance with the balance beam. The children will provide their answers by talking to the robot. The input of the state of the learning material for the robot will be controlled by a 'Wizard of Oz' technique.

In the control condition learners will be working on the same assignment but without the robot. In this case the tablet provides instruction and will pose exactly the same questions to help learners explore the phenomenon of balance. In the control condition there is no background story, but children are asked to do the assignment as part of their educational program. The children will provide their answers verbally, and it will seem as if the tablet records the answers. In both conditions video recordings will be made of the children working on the task. It is hypothesized that when working on the task in an appropriate social context, in this case being accompanied by the robot, giving answers to the questions will result in more verbal explanatory behavior. Verbally explaining to another person can facilitate greater understanding of one's own ideas and knowledge [23] and might therefore lead to better learning and transfer [36].

The second experiment will focus on the expected cognitive competence children believe the robot has. There will be three conditions. In all conditions the robot will make some incorrect suggestions. The difference between the conditions is that the children are primed to believe that the robot is (1) an expert, (2) a novice or (3) no priming. The goal is to find out how competent and trustworthy children believe the robot is before and after the experiment.

In this paper, we have described some aspects of an initial theoretical framework that we use to design our experiments and user studies to investigate child-robot symbiotic interaction. We are going to give an emphasis to the process of learning in different contexts, focusing on collaborative learning and exploiting the robot as an adaptive co-learner. Thus the robot can scaffold the child to go through an effective learning process. For the future work we aim to investigate how a social robot can scaffold child's inquiry process by facilitating the expansion of ZPD in an effective and enjoyable way focusing on the development of children's meta-cognitive skills.

ACKNOWLEDGMENT

This project has received funding from the European Union Seventh Framework Programme (FP7-ICT-2013-10) as part of EASEL under grant agreement n° 611971.

REFERENCES

[1] Douglas, A.E., *The Symbiotic Habit*. Princeton: Princeton University Press. (2010).

[2] Foston, C.T. and R.S. Perry, *Constructivism: A psychological Theory of Learning*, in *Constructivism*, C.T. Foston, Editor., Teachers College, Columbia University: New York. (2005).

[3] Piaget, J. and B. Inhelder, *The psychology of the child*. New York: Basic Books. (1969).

[4] Piaget, J., *The theory of stages in cognitive development*, in *Measurement and Piaget*, D.R. Green, Editor., McGraw-Hill: New York. p. 1-11. (1971).

[5] Papert, S., *Papert on Piaget*, in *Time: The Century's Greatest Minds*. p. 105. (1999).

[6] Donaldson, M., *Children's minds*. Glasgow: Fontana. (1978).

[7] Smith, L.B. and E. Thelen, *Development as a dynamic system.* Trends in Cognitive Sciences, **7**(8): p. 343-348, (2003).

[8] Resnick, M. and E. Rosenbaum, *Designing for tinkerability*, in *Design, make, play: growing the next generation of STEM Innovators*, M. Honey and D. Kanter, Editors., Routledge: New York. p. 163-181. (2013).

[9] Banchi, H. and R.L. Bell, *The many levels of inquiry.* Science and Children, **46**: p. 26-29, (2008).

[10] Pellegrini, A.D., *The Role of Play in Human Development*. New York: Oxford University Press. (2009).

[11] Whitebread, D. and D. Pino Paternak, *Metacognition, self-regulation and meta-knowing*, in *International Handbook of Psychology in Education*, K. Littleton, C. Wood, and J. Kleine Staarman, Editors., Emerald: Bingley, UK. (2010).

[12] NSF, *An introduction to inquiry*, in *Inquiry: Thoughts, Views and Strategies for K-5 Classroom*. National Science Foundation: Washington. p. 1-5. (2000).

[13] van Joolingen, W.R., T. de Jong, and A. Dimitrakopoulou, *Issues in computer supported inquiry learning in science.* Journal of Computer Assisted Learning, **23**: p. 111-119, (2007).

[14] Klahr, D., *Exploring science: the cognition and development of discovery processes*. Cambridge: The MIT Press. (2000).

[15] Klahr, D. and K. Dunbar, *Dual search space during scientific reasoning.* Cognitive Science, **12**: p. 1-48, (1988).

[16] Bell, R.L., L. Smetana, and I. Binns, *Simplifying inquiry instruction.* The Science Teacher, **72**: p. 30-33, (2005).

[17] Vygotsky, L.S., *Mind in Society: The Development of Higher Psychological Process*. Cambridge, MA: Harvard University Press. (1930-34/1978).

[18] Corsaro, W., *The sociology of childhood*. Thousand Oaks, California: Pine Forge Press. (1997).

[19] Vygotsky, L.S., *Play and its role in the development of the child.* Soviet Psychology, **12**(6): p. 62-76, (1933/1966).

[20] Siegler, R.S., *Microgenetic Analyses of Learning*, in *Handbook of Child Psychology*, D. Kuhn and R.S. Siegler, Editors., Wiley: Hoboken, NJ. p. 464-510. (2006).

[21] Rogoff, B., *Cognition as a Collaborative Process*, in *Handbook of Child Psychology*, D. Kuhn and R.S. Siegler, Editors., Wiley: New York. p. 679-744. (1998).

[22] Chi, M.T.H., et al., *Self-explanations: how students study and use examples in learning to solve problems.* Cognitive Science, **13**: p. 145-182, (1989).

[23] Holmes, J., *Designing agents to support learning by explaining.* Computers and Education, **48**: p. 523-547, (2007).

[24] Ploetzner, R., et al., *Learning by explaining to oneself and to others*, in *Collaborative learning: Cognitive and Computational Approaches*, P. Dillenbourg, Editor., Elsevier: Oxford. p. 103-121. (1999).

[25] Wood, D., J.S. Bruner, and G. Ross, *The role of tutoring in problem solving.* Journal of Child Psychology and Psychiatry, **17**(2): p. 89-100, (1976).

[26] Topping, K.J., *Trends in Peer Learning.* Educational Psychology, **25**(6): p. 631-645, (2005).

[27] Kort, B., R. Reilly, and R.W. Picard, *An affective model of interplay between emotions and learning: reenfineering educational pedagogy - Building a learning companion*, in *International Conference of Advanced Learning Technologies*. IEEE Computer Society: Washington, DC. (2001).

[28] Fredrickson, B.L., *The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions.* American Psychologist, **56**(3): p. 218-226, (2001).

[29] Dolcos, F., A.D. Iordan, and S. Dolcos, *Neural correlates of emotion-cognition interactions: Areview of evidence from brain imaging investigations.* Journal of Cognitive Psychology, **23**(6): p. 669-694, (2011).

[30] Klein, G. and P. Feltovich, *Common Ground and Coordination in Joint Activity.* Organizational simulation: p. 1-42, (2005).

[31] Clark, H., *Using Language.* (1996).

[32] Simpson, J.A., *Psychological Foundations of Trust.* Current Directions in Psychological Science, **16**: p. 264-268, (2007).

[33] Lee, J.D. and K.A. See, *Trust in Automation: Designing for Appropriate Reliance.* Human Factors: The Journal of the Human Factors and Ergonomics Society, **46**: p. 50-80, (2004).

[34] Kanda, T., et al., *Interactive robots as social partners and peer tutors for children: a field trial.* Human-Computer Interaction, **19**(1): p. 61-84, (2004).

[35] Belpaeme, T., et al., *Multimodal Child-Robot Interaction: Buiding Social Bonds.* Journal of Human-Robot Interaction, **1**(2): p. 33-53, (2012).

[36] Rittle-Johnson, B., M. Saylor, and K. Swygert, *Learning from explaining: does it matter if mom is listening?* Journal of Child Psychology, **100**(3): p. 215-224, (2008).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Does anyone want to talk to me? – Reflections on the use of assistance and companion robots in care homes

Kerstin Dautenhahn[1] and **Anne Campbell**[2] and **Dag Sverre Syrdal**[1]

1 Introduction: Robotic Companions for Elderly People

A growing number of research efforts worldwide aim at developing assistive robots to help elderly people in their own homes or in care homes. The rationale for home assistance robotic technology is based on demographic changes in many countries worldwide, with an ageing population. For example, it is predicted that in the European Union the number of people over 65 years will almost double (by 2060) and the number of people between 15-64 years will decrease by over 10%. Health care costs are also rising [33]. Developments into home companions and solutions for Ambient Assisted Living (AAL) in elderly peoples homes or care homes have grown significantly in the EU, see projects such as SRS[12] , Hermes[5] , Florence [4] , KSERA[7] , MOBISERV [9] , Rubicon [11], ACCOMPANY [1] or ROBOT-ERA[10] , to name a few. Recent videos of results on smart home companion robots and the type of assistance they can provide have been illustrated for MOBISERV[29] and ACCOMPANY[15] . Products for robots used in peoples homes are beginning to be marketed, cf. Toyota's Human Support Robot (HSR)[13], Mitsubishi's communication robot Wakamaru[14] , Aldebaran's Pepper robot [2], or Cynthia Breazeal's Jibo robot[6]. These robots come in different shapes and sizes, and appearance and behaviour will influence which roles these robots are being assigned to by their users and the human-robot relationships that may emerge.

One of the authors has been involved in European projects on home assistance robots since 2004, as part of the COGNIRON [3] , LIREC [8] and ACCOMPANY [1] projects. COGNIRON was one of the first projects in Europe on home companion robots. One lesson learnt during the project was the need to move out of the laboratory and into a realistic home setting, which led to the acquisition and development of the University of Hertfordshire Robot House, a smart home equipped with a sensor network and robots being able to detect daily living activities and provide physical, social and cognitive assistance. A second lesson was the need to move away from Wizard-of-Oz (remote controlled) studies. In LIREC the emphasis was on developing fully autonomous home assistant robots, with an emphasis on social assistance. During ACCOMPANY, this direction has been elaborated and extended through allowing the robot to be taught and shown new behaviours and routines by the user, including evaluations with elderly users and their formal and informal carers in long-term studies in three European countries. The ACCOMPANY project has particularly advanced a direction where such autonomously operating companion robots, as part of a smart home infrastructure, socially engage and assist the user, using personalization and human-robot teaching and co-learning for reablement of the user[36] . While these projects have focused primarily on the use of robot within home settings, a separate strand of research within the University of Hertfordshire's work in ACCOMPANY actively elicited the views of residents and staff at a local care home, through the use of theatre prototyping[34] followed by interviews as previously reported in Walters et al.[45]. The current position paper draws on these experiences and findings, as well as those from the other projects, to consider the role that social robots may play in a care home environment.

[1] School of Computer Science, University of Hertfordshire, email: {k.dautenhahn; d.s.syrdal} @herts.ac.uk
[2] School of Health and Social Work, University of Hertfordshire, email a.2.campbell@herts.ac.uk

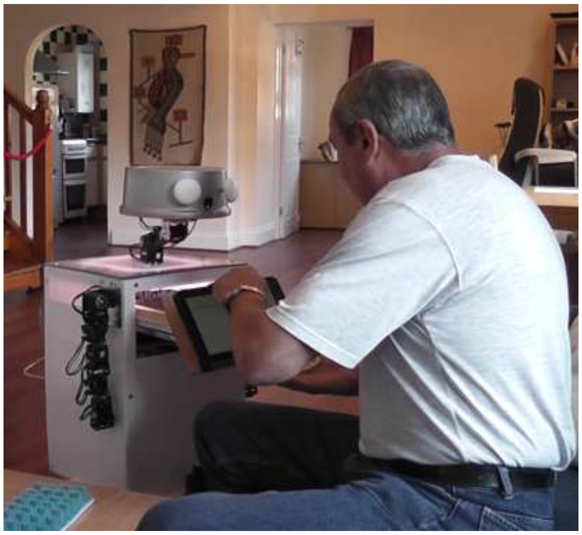
2 Roles of Robots

Different roles of robots in human society have been proposed[21] , including a machine operating without human contact; a tool in the hands of a human operator; a peer as a member of a humaninhabited environment; a robot as a persuasive machine influencing people's views and/or behaviour (e.g. in a therapeutic context); a robot as a social mediator mediating interactions between people; a robot as a model social actor. Opinions on viewing robots either as friends, assistants or butlers have been investigated [23] . It has been suggested the robot can act as a mentor for humans, or information consumer whereby a human uses information provided by a robot[25] . Further roles that have been introduced view robots as a team member in collaborative tasks [19] or roles for robots as learners [39, 28]. Companion robots have been defined as robots that not only can carry out a range of useful tasks, but do so in a socially acceptable manner [22] . This role typically involves both long-term and repeated interaction, as is the case for robots used in an elderly person's home or in a care home. Will people develop human-like relationships with such companion robots? Some studies have tried to address these from a user-centric point of view. Beer et al.[18] found that participants primarily focused on the ability of the robot to streamline and reduce the amount of effort required to maintain their household. However, a recent study based on both recent literature research and focus groups with 41 elderly people, 40 formal caregivers and 32 informal caregivers in the Netherlands, UK and France, the most problematic challenges to independent living were identified mobility, self-care, and interpersonal interaction and relationships [17].

Thus, there seem to be two domains where robots are envisaged to assist in: the physical and/or cognitive domain, providing e.g specific assistance in remembering events and appointments, or to move around, and the domain of social relationships.

This duality of roles do exist in how robots are being proposed to be used in such settings, while surveys of envisaged use scenarios

Figure 1. A companion robot at the University of Hertfordshire



indicate that medical and healthcare personnel see robots as tools that can provide physical assistance with their tasks [40] , however, there are also studies investigating the value of robots as companions in these settings[38].

This approach is grounded in that, apart from physical needs, a key problem in care homes is the resident's loneliness. It impacts upon ' quality of life and wellbeing, adversely affects health and increases the use of health and social care services . A number of interventions have been used, e.g. one-to-one approaches such as Befriending, Mentoring, group services such as lunch clubs, or community engagement through public facilities (sports etc) [46]. Interestingly, in a recent approach chickens have been introduced to a care home, and proved popular with both staff and residents[35]. The impact of robots and animals can be directly compared[16]. Could robots become part of such services?

3 Ethical Issues

While this short position paper cannot comprehensively address the ethical issues involved in the adoption of robots in elder care and the associated literature, we note that elsewhere the danger to anthropomorphise and romanticise robots has been highlighted[20]. The roles that are ascribed to robots and the human—robot relationships discussed in the research community are predominantly based on terms that originally describe human-human interactions. So there is a tendency to use terms robotic 'assistant' or robotic 'carer' and apply the human equivalent literally which automatically implies a whole range of different human-like qualities and abilities, that robots at present cannot address, in terms of their physical and cognitive abilities, as well as in terms of their emotional intelligence, as well as ethical and moral judgements. A number of ethical considerations need to be considered when fostering social relationships between robots and elderly people. Sherry Turkle[41] has previously discussed the danger of 'relational artifacts', i.e. robot designed specifically to encourage people to form a relationship with them. She argued that such 'non-authentic' interaction may lead to people preferring the (relatively easy and predictable and non-judgemental) interaction with a robot compared to interactions with real people. Specifically with regard to eldercare, Amanda and Noel Sharkey[37] pointed out risks involved in using robots in elder care, including the potential for the reduction in the amount of human contact as well as concerns about deception and infantilisation. The theme of deception, infantilisation and the possible reduction in human contact is also emphasized in other reflections on ethical norms of using robots in caring role for elderly people[42, 24].

Interestingly, designing robots as interactive systems that people can engage with, e.g. play games with, is technically feasible. Even pet-like, non-humanoid robots such as Paro have been shown to be successful companions[30]. On the other hand, providing physical assistance involves many technical challenges e.g. in terms of object manipulation, navigation, safety, etc. Thus, if it is 'easier' to build robots as socially interactive companions, and to focus on its role to engage people, shall one concentrate research efforts on this aspect? Is it ethically justifiable, desirable and acceptable by elderly people and their carers, given the above mentioned concerns of deception, infantilisation, and providing non-authentic experiences? In order to shed some initial light on these issues, one of the authors conducted interviews in a care home for elderly people.

4 INTERVIEWS STUDY WITH RESIDENTS AND CARER IN A CARE HOME

An interview study was conducted with carers and residents of a care home in UK. In this study, residents and staff at the residential care home were shown a play which focused on how the adoption of personal home companion impacted the relationships in a domestic household. The play and other aspects of the study is briefly summarised here, details are provided elsewhere[45] . While the play focused on the use of a robot in a different environment, it served to raise awareness of how robots may assist in, and influence the daily life of their users. We would also note that there was no verbal interaction from the robot in the play. Three months after the play, a follow-up study was conducted in which three residents, all with learning disabilities and/or physical disabilities were interviewed, followed by interviews of three experienced registered nurses. The 15-20 min interviews took place in the communal dining room of the home that is familiar and comfortable to both residents and carers. A semi-structured interview technique was used since it is considered a reliable and flexible method and can cater for some of the residents' disabilities[32]. The interviewer wrote down the interview data during the interview, an approach considered less intrusive than audio-taping the interviews. Based on these notes, the interviewer conducted a content analysis of the interview data a number of themes emerged that are described in detail in Walters et al. [45]. Relevant for the present article are the following themes and comments from residents and carers: Concerning acceptable boundaries for care by humans and robots, one resident said that the most important care for her from the robot was psychological care:

> *'Make me feel lovely in myself and give me a boost...make things different...I want to dance with it'.*
> *'I would like the robot to be chatty and to nod his head to show he has heard me'.*

Two other residents wanted the robot to *'Tidy my room and maybe feed me in the future'* and *'comb my hair'*. Regarding conversation and companionship, one of the interviewed residents wanted the robot to be able to start a conversation and then acknowledge that he had heard about her sore knee. Another wanted the robot to dance with her. One theme arising from the interviews of the registered nurses concerned how the robot could provide assistance to staff and residents, while they still preferred a human to a robot colleague. All 3 nurses thought the robots would help with both physical and psychological care:

> *'They could provide company, socialise and boost morale'.*
> *'They could be friendly, shake hands and make friendly sounds; talk to them and reduce loneliness'.*
> *'Help with feeding and walking beside them would be helpful'.*

Concerning conversation and companionship all three nurses would really value robot that can engage in conversations with residents and provide stimulation:

> *'Stimulation helps residents feel important'.*
> *'Helpful when staff are busy'.*

5 Reflections

The interview study above highlighted a number of issues in favour of robot providing social interaction and communication with residents in a care home in order to help with their loneliness. There are also a number of practical issues, based on experience gained by the second author in care homes, that would support robots in that role:

- The group of residents in care homes is often diverse, ranging from people with dementia, people with learning disabilities, people terminally ill e.g. with cancer, and others. This diversity can impact on the willingness and enjoyment of residents to talk to each other
- Residents in a care home do not know each other prior to joining the care home, they are not a naturally formed unit of friends or family. We cannot expect randomly created groups of people to make friends easily, or even to be interested in talking to each other, while having to live under the same roof under a daily basis.
- Care staff is often very focused on task and efficiency, often under a lot of time-pressure to 'get things done'. There is a large spectrum in the quality of care, but in some care homes social interaction with residents might not be high on the priority list of care staff and their managers.
- From the point of view of care staff, interaction with residents may not always be as enjoyable as one might envisage, e.g. due to memory problems people with dementia may engage in very repetitive conversations.
- In a social environment such as a care home, residents might feel not 'getting along with the others', due to real or perceived conflicts with other residents.
- Some residents may have psychiatric conditions which make them feel paranoid and sometimes aggressive.
- Care home staff and/or residents may not all have English as their first language which affects their ability to communicate with each other smoothly. There may also be differences in intercultural understanding of what is socially acceptable conversation.

Thus, while in an ideal world, care homes should be places where carers and residents live together as 'one happy family', the reality often differs. And it may be useful for robots to provide opportunities for communication and interaction, even if interaction with robots is mechanical, and lacks authenticity and depths of human contact as we have argued elsewhere[41, 22] . For example, present robots cannot replace the gentleness and meaningfulness of a person stroking someone's hair, or touching someone's hands, or a comforting word. This does not always mean that the robot will have to replace carer-resident or resident-resident interactions. Rather, it may function as a social facilitator, or mediator, and may be able to assist residents and carers in overcoming some of the practical issues that often restrict human-human interactions in care homes. Previous research has suggested that the presence of a robot in a care may work to facilitate a greater degree of interaction between the residents of the care home [27, 43] , and this effect may be leveraged further by using features like a memory visualisation system (which uses photos and text to create narratives of previous interaction)[26] to aid further when trying creating common ground between human interactants . In addition, there is also the possibility to adapt and apply research in using robots to increase dyadic interactions in other user-groups [44, 31] in order to further the ability of a robot companion as a social facilitator or mediator. While it can be argued that some of the issues, in particular the staff's focus on task and efficiency can be mediated by the adoption of robots to provide physical support with some of the tasks, this does not necessarily address the other points raised here. We do not argue for robots to replace carers or human contact in general, however, we argue that in situations where residents can expect, and may suffer from, only very little human contact that in such circumstances robots could be beneficial to them and their carers, by helping them to feel less lonely, not only through the direct interaction between the resident and the robot, but also through the robot's ability to mediate interactions between residents and residents and carers — and thus improving the health and well-being of the residents as well as the working conditions and atmosphere at work as experienced by the staff.

ACKNOWLEDGEMENTS

We would like to thank the referees for their comments which helped improve this paper.

REFERENCES

[1] ACCOMPANY. http://rehabilitationrobotics.net/cms2/.
[2] Aldebaran Pepper. https://www.aldebaran.com/en/a-robots/who-is-pepper.
[3] Cogniron. http://www.cogniron.org/final/Home.php.
[4] Florence. http://florence-project.eu.
[5] Hermes. http://fp7-hermes.eu.
[6] Jibo. http://www.myjibo.com/.
[7] KSERA. http://ksera.ieis.tue.nl.
[8] LIREC. http://lirec.eu/project.
[9] MOBISERV. http://www.mobiserv.info/.
[10] Robot-ERA. http://robot-era.eu.
[11] Rubicon. http://fp7rubicon.eu/.
[12] SRS. http://srs-project.eu.
[13] Toyota Human-Support Robot. http://www.toyota-global.com/innovation/partner_robot/.
[14] Wakamaru. https://www.mhi-global.com/products/detail/wakamaru_about.html.
[15] ACCOMPANY. ACCOMPANY Project Video. http://www.youtube.com/watch?v=Z1MJPdhniXc.
[16] Marian R Banks, Lisa M Willoughby, and William A Banks, 'Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs', *Journal of the American Medical Directors Association*, **9**(3), 173–177, (2008).
[17] Sandra Bedaf, Gert Jan Gelderblom, Dag Sverre Syrdal, Hagen Lehmann, Hervé Michel, David Hewson, Farshid Amirabdollahian, Kerstin Dautenhahn, and Luc de Witte, 'Which activities threaten independent living of elderly when becoming problematic: inspiration for meaningful service robot functionality', *Disability and Rehabilitation: Assistive Technology*, **9**(6), 445–452, (2013).
[18] Jenay M Beer, Cory-Ann Smarr, Tiffany L Chen, Akanksha Prakash, Tracy L Mitzner, Charles C Kemp, and Wendy A Rogers, 'The domesticated robot: design guidelines for assisting older adults to age in

place', in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pp. 335–342. ACM, (2012).
[19] Cynthia Breazeal, Andrew Brooks, Jesse Gray, Guy Hoffman, Cory Kidd, Hans Lee, Jeff Lieberman, Andrea Lockerd, and David Chilongo, 'Tutelage and collaboration for humanoid robots', *International Journal of Humanoid Robotics*, **1**(02), 315–348, (2004).
[20] Kerstin Dautenhahn, 'Human-robot interaction', in *The Encyclopedia of Human-Computer Interaction, 2nd Ed*, eds., Mads Soegaard and Rikke Friis Dam.
[21] Kerstin Dautenhahn, 'Roles and functions of robots in human society: implications from research in autism therapy', *Robotica*, **21**(04), 443–452, (2003).
[22] Kerstin Dautenhahn, 'Socially intelligent robots: dimensions of human–robot interaction', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **362**(1480), 679–704, (2007).
[23] Kerstin Dautenhahn, Sarah Woods, Christina Kaouri, Michael L Walters, Kheng Lee Koay, and Iain Werry, 'What is a robot companion-friend, assistant or butler?', in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pp. 1192–1197. IEEE, (2005).
[24] Heather Draper, Tom Sorell, Sandra Bedaf3 Christina Gutierrez Ruiz, Hagen Lehmann, Michael Hervé, Gert Jan Gelderblom, Kerstin Dautenhahn, and Farshid Amirabdollahian, 'What asking potential users about ethical values adds to our understanding of an ethical framework for social robots for older people.', *MEMCA-14. This Proceedings*, (2014).
[25] Michael A Goodrich and Alan C Schultz, 'Human-robot interaction: a survey', *Foundations and trends in human-computer interaction*, **1**(3), 203–275, (2007).
[26] Wan Ching Ho, Kerstin Dautenhahn, Nathan Burke, Joe Saunders, and Joan Saez-Pons, 'Episodic memory visualization in robot companions providing a memory prosthesis for elderly users', *Assistive Technology*, (2013).
[27] Cory D Kidd, Will Taggart, and Sherry Turkle, 'A sociable robot to encourage social interaction among the elderly', in *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pp. 3972–3976. IEEE, (2006).
[28] Katrin S Lohan, Karola Pitsch, Katharina J Rohlfing, Kerstin Fischer, Joe Saunders, Hagen Lehmann, Chrystopher Nehaniv, and Britta Wrede, 'Contingency allows the robot to spot the tutor and to learn from interaction', in *Development and Learning (ICDL), 2011 IEEE International Conference on*, volume 2, pp. 1–8. IEEE, (2011).
[29] MOBISERV. MOBISERV Project Video. `http://www.youtube.com/watch?feature=player_embedded&v=6DFJwnwzhPs`.
[30] Wendy Moyle, Marie Cooke, Elizabeth Beattie, Cindy Jones, Barbara Klein, Glenda Cook, and Chrystal Gray, 'Exploring the effect of companion robots on emotional expression in older adults with dementia: a pilot randomized controlled trial.', *Journal of gerontological nursing*, **39**(5), 46–53, (2013).
[31] Fotios Papadopoulos, Kerstin Dautenhahn, and Wan Ching Ho, 'Exploring the use of robots as social mediators in a remote human-human collaborative communication experiment', *Paladyn*, **3**(1), 1–10, (2012).
[32] Denise F Polit and Cheryl Tatano Beck, *Nursing research: Principles and methods*, Lippincott Williams & Wilkins, 2004.
[33] Bartosz Przywara, 'Projecting future health care expenditure at european level: drivers, methodology and main results', Technical report, Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission, (2010).
[34] Mark Rice, Alan Newell, and MAGGIE Morgan, 'Forum theatre as a requirements gathering methodology in the design of a home telecommunication system for older adults', *Behaviour & Information Technology*, **26**(4), 323–331, (2007).
[35] Jessica Salter. Chickens helping the elderly tackle loneliness. `http://www.telegraph.co.uk/news/health/11198410/Chickens-helping-the-elderly-tackle-loneliness.html`.
[36] Joe Saunders, Nathan Burke, Kheng Lee Koay, and Kerstin Dautenhahn, 'A user friendly robot architecture for re-ablement and co-learning in a sensorised homes', in *Assistive Technology: From Research to Practice: AAATE 2013*, eds., Pedro Encarnacao, Luis Azevedo, and Gert Jan Gelderblom, volume 33, IOS Press, (2013).
[37] Amanda Sharkey and Noel Sharkey, 'Granny and the robots: ethical issues in robot care for the elderly', *Ethics and Information Technology*, **14**(1), 27–40, (2012).
[38] Will Taggart, Sherry Turkle, and Cory D Kidd, 'An interactive robot in a nursing home: Preliminary remarks', in *Towards Social Mechanisms of Android Science: A COGSCI Workshop*, (2005).
[39] Andrea L Thomaz and Cynthia Breazeal, 'Teachable robots: Understanding human teaching behavior to build more effective robot learners', *Artificial Intelligence*, **172**(6), 716–737, (2008).
[40] Katherine M Tsui and Holly A Yanco, 'Assistive, rehabilitation, and surgical robots from the perspective of medical and healthcare professionals', in *AAAI 2007 Workshop on Human Implications of Human-Robot Interaction, Technical Report WS-07-07 Papers from the AAAI 2007 Workshop on Human Implications of HRI*, (2007).
[41] Sherry Turkle, 'Authenticity in the age of digital companions', *Interaction Studies*, **8**(3), 501–517, (2007).
[42] Shannon Vallor, 'Carebots and caregivers: Sustaining the ethical ideal of care in the twenty-first century', *Philosophy & Technology*, **24**(3), 251–268, (2011).
[43] Kazuyoshi Wada and Takanori Shibata, 'Robot therapy in a care house-results of case studies', in *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pp. 581–586. IEEE, (2006).
[44] Joshua Wainer, Ben Robins, Farshid Amirabdollahian, and Kerstin Dautenhahn, 'Using the humanoid robot kaspar to autonomously play triadic games and facilitate collaborative play among children with autism', *Autonomous Mental Development, IEEE Transactions on*, **6**(3), 183–199, (2014).
[45] Michael L Walters, Kheng Lee Koay, Dag Sverre Syrdal, Anne Campbell, and Kerstin Dautenhahn, 'Companion robots for elderly people: Using theatre to investigate potential users' views', in *RO-MAN, 2013 IEEE*, pp. 691–696. IEEE, (2013).
[46] Karen Windle, Karen Francis, and Caroline Coomber. SCIE Research briefing 39: Preventing loneliness and social isolation: interventions and outcomes. `http://www.scie.org.uk/publications/briefings/briefing39/index.asp`.

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

Robots Have Needs Too: People Adapt Their Proxemic Preferences to Improve Autonomous Robot Recognition of Human Social Signals

Ross Mead[1] and **Maja J Matarić**[2]

Abstract. An objective of autonomous socially assistive robots is to meet the needs and preferences of human users. However, this can sometimes be at the expense of the robot's own ability to understand *social signals* produced by the user. In particular, human preferences of distance (*proxemics*) to the robot can have significant impact on the performance rates of its automated speech and gesture recognition systems. In this work, we investigated how user proxemic preferences changed to improve the robot's understanding human social signals. We performed an experiment in which a robot's ability to understand social signals was artificially varied, either *uniformly* or *attenuated* across distance. Participants ($N = 100$) instructed a robot using speech and pointing gestures, and provided their proxemic preferences before and after the interaction. We report two major findings: 1) people predictably underestimate (based on a Power Law) the distance to the location of robot peak performance; and 2) people adjust their proxemic preferences to be near the *perceived* location of robot peak performance. This work offers insights into the dynamic nature of human-robot proxemics, and has significant implications for the design of social robots and robust autonomous robot proxemic control systems.

1 Introduction

A social robot utilizes natural communication mechanisms, such as speech and gesture, to autonomously interact with humans to accomplish some individual or joint task [2]. The growing field of socially assistive robotics (SAR) is at the intersection of social robotics and assistive robotics that focuses on non-contact human-robot interaction (HRI) aimed at monitoring, coaching, teaching, training, and rehabilitation domains [4]. Notable areas of SAR include robotics for older adults, children with autism spectrum disorders, and people in post-stroke rehabilitation, among others [25, 17].

Consequently, SAR constitutes an important subfield of robotics with significant potential to improve health and quality of life. Because the majority of SAR contexts investigated to date involve one-on-one face-to-face interaction between the robot and the user, how the robot understands and responds to the user is crucial to successful autonomous social robots [1], in SAR contexts and beyond.

One of the most fundamental social behaviors is *proxemics*, the social use of space in face-to-face social encounters [5]. A mobile social robot must position itself appropriately when interacting with the user. However, robot position has a significant impact on the robot's *performance*—in this work, performance is measured by automated speech and gesture recognition rates. Just like electrical signals, human *social signals* (e.g., speech and gesture) are *attenuated* (lose signal strength) based on distance, which dramatically changes the way in which automated recognition systems detect and identify the signal; thus, a proxemic control system that often varies its location and, thus, creates signal attenuation, can be a defining factor in the success or failure of a social robot [16].

In our previous work [16] (described in detail in Section 2.2), we modeled social robot performance attenuated by distance, which was then used to implement an autonomous robot proxemic controller that maximizes its performance during face-to-face HRI; however, this work begged the question as to whether or not people would accept a social robot that positions itself in a way that differs from traditional user proxemic preferences. Would users naturally change their proxemic preferences if they observed differences in robot performance in different proxemic configurations, or would their proxemic preferences persist, mandating that robot developers must improve autonomous speech and gesture recognition systems before social and socially assistive robot technology can be deployed in the real world? This question is the focus of the investigation reported here.

2 Background

The anthropologist Edward T. Hall [5] coined the term "proxemics", and, in [6], proposed that proxemics lends itself well to being analyzed with performance (as measured through sensory experience) in mind. Proxemics has been studied in a variety of ways in HRI; here, we constrain our review of related work to that of *autonomous* HRI[3].

2.1 Comfort-based Proxemics in HRI

The majority of proxemics work in HRI focuses on maximizing user *comfort* during a face-to-face interaction. The results of many human-robot proxemics studies have been consolidated and normalized in [28], reporting mean distances of 0.49–0.71 meters using a variety of robots and conditions. Comfort-based proxemic preferences between humans and the PR2 robot[4] were investigated in [24], reporting mean distances of 0.25–0.52 meters; in [16], we investigated the same preferences using the PR2 in a conversational context, reporting a mean distance of 0.94 meters. Farther proxemic preferences have been measured in [18] and [26], reporting mean distances of 1.0–1.1 meters and 1.7–1.8 meters, respectively.

[1] University of Southern California, USA, email: rossmead@usc.edu
[2] University of Southern California, USA, email: mataric@usc.edu

[3] There is a myriad of related work reporting how humans adapt to various technologies, but this is beyond the scope of this work. For a review, see [8].
[4] https://www.willowgarage.com/pages/pr2/overview

However, results in our previous work [16] suggest that autonomous speech and gesture recognition systems do not perform well using comfort-based proxemic configurations. Speech recognition performed adequately at distances less than 2.5 meters, and face and hand gesture recognition performed adequately at distances of 1.5–2.5 meters; thus, given current technologies, distances for mutual recognition of these social signals is between 1.5 and 2.5 meters, at and beyond the far end of comfort-based proxemic preferences.

2.2 Performance-based Proxemics in HRI

Our previous work utilized advancements in markerless motion capture (specifically, the Microsoft Kinect) to automatically extract proxemic features based on metrics from the social sciences [11, 14]. These features were then used to recognize spatiotemporal interaction behaviors, such as the initiation, acceptance, aversion, and termination of an interaction [12, 14]. These investigations offered insights into the development of proxemic controllers for autonomous social robots, and suggested an alternative approach to the representation of proxemic behavior that goes beyond simple distance and orientation [13]. A probabilistic framework for autonomous proxemic control was proposed in [15, 10] that considers *performance* by maximizing the sensory experience of each agent (human or robot) in a co-present social encounter. The methodology established an elegant connection between previous approaches and illuminated the functional aspects of proxemic behavior in HRI [13], specifically, the impact of spacing on speech and gesture behavior recognition and production. In [16], we formally modeled (using a dynamic Bayesian network [9]) autonomous speech and gesture recognition systems as a function of distance and orientation between a social robot and a human user, and implemented the model as an autonomous proxemic controller, which was shown to maximize robot performance in HRI.

However, while our approach to proxemic control *objectively* maximized the performance of the robot, it also resulted in proxemic configurations that are atypical for human-robot interactions (e.g., positioning itself farther or nearer to the user than preferred). Thus, the question arose as to whether or not people would *subjectively* adopt a technology that places performance over preference, as it might place a burden on people to change their own behaviors to make the technology function adequately.

2.3 Challenges in Human Spatial Adaptation

For humans to adapt their proxemic preferences to a robot, they must be able to accurately identify regions in which the robot is performing well; however, errors in human distance estimation increase non-linearly with increases in distance, time, and uncertainty [19]. Fortunately, the relationship between human distance estimation and each of these factors is very well represented by Steven's Power Law, ax^b, where x is distance [19, 23]. Unfortunately, these relationships are reported for distances of 3–23 meters, which are farther away than in those with which we are concerned for face-to-face HRI—thus, we cannot use the reported model parameters and must derive our own.

In this work, we investigate how user proxemic preferences change in the presence of a social robot that is recognizing and responding to instructions provided by a human user. Robot performance (ability to understand speech and gesture) is artificially attenuated to expose participants to success and failure scenarios while interacting with the robot. In Section 3, we describe the overall setup in which our investigation took place. In Section 4, we outline the specific procedures, conditions, hypotheses, and participants of our experiment.

3 Experimental Setup

3.1 Materials

The experimental robotic system used in this work was the Bandit upper-body humanoid robot[5] [Figure 1]. Bandit has 19 degrees of freedom: 7 in each arm (shoulder forward-and-backward, shoulder in-and-out, elbow tilt, elbow twist, wrist twist, wrist tilt, grabber open-and-close; left and right arms), 2 in the head (pan and tilt), 2 in the lips (upper and lower), and 1 in the eyebrows. These degrees of freedom allow Bandit to be expressive using individual and combined motions of the head, face, and arms. Mounted atop a Pioneer 3-AT mobile base[6], the entire robot system is 1.3 meters tall.

A Bluetooth PlayStation 3 (PS3) controller served as a remote control interface with the robot. The controller was used by the experimenter (seated behind a one-way mirror [Figure 2]) to step the robot through each part of the experimental procedure (described in Section 4.1)—the decisions and actions taken by the robot during the experiment were completely autonomous, but the timing of its actions were controlled by the press of a "next" button. The controller was also used to record distance measurements during the experiment, and to provide ground-truth information to the robot as to what the participant was communicating (however, the robot autonomously determined how to respond based on the experimental conditions described in Section 4.2).

Four small boxes were placed in the room, located at 0.75 meters and 1.5 meters from the centerline on each side (left and right) of the participant [Figure 2]. During the experiment (described in Section 4.1), the participant instructed the robot to look at these boxes. Each box was labeled with a unique shape and color; in this experiment, the shapes and colors matched the buttons on the PS3 controller: a green triangle, a red circle, a blue cross, and a purple square. This allowed the experimenter to easily indicate to the robot to which box the user was attending (i.e., "ground-truth").

A laser rangefinder on-board the robot was used to measure the distance from the robot to the participant's legs at all times.

Figure 1. The Bandit upper-body humanoid robot.

[5]http://robotics.usc.edu/interaction/?l=Laboratory:Robots#BanditII
[6]http://www.mobilerobots.com/ResearchRobots/P3AT.aspx

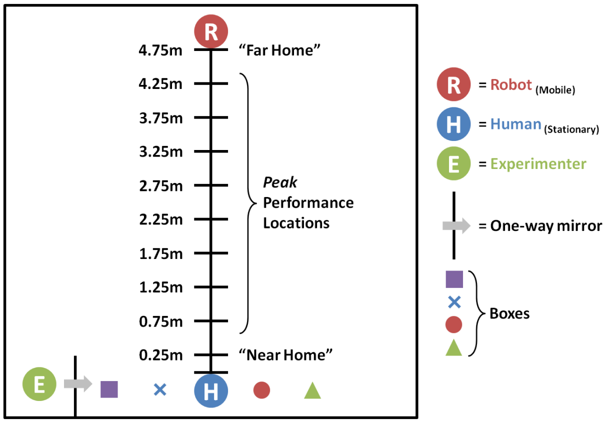


Figure 2. The experimental setup.

3.2 Robot Behaviors

The robot autonomously executed three primary behaviors throughout the experiment: 1) forward and backward base movement, 2) maintaining eye contact with the participant, and 3) responding to participant instructions with head movements and audio cues.

Robot base movement was along a straight-line path directly in front of the participant, and was limited to distances of 0.25 meters (referred to as the “near home” location) and 4.75 meters (referred to as the “far home” location); it returned repeatedly to these “home” locations throughout the experiment. Robot velocity was proportional to the distance to the goal location; the maximum robot speed was 0.3 m/s, which people find acceptable [22].

As the robot moved, it maintained eye contact with the participant. The robot has eyes, but they are not actuated, so the robot’s head pitched up or down depending on the location of the participant’s head, which was determined by the distance to the participant (from the on-board laser) and the participant’s self-reported height. We note that prolonged eye contact from the robot often results in user preferences of increased distance in HRI [24, 18].

The robot provided head movement and audio cues to indicate whether or not it understood instructions provided by the participant (described in Section 4.1.2). If the robot understood the instructions, it provided an *affirmative response* (looking at a box); if the robot did not understand the instructions, it provided a *negative response* (shaking its head). With each head movement, one of two affective sounds were also played to supplement the robot’s response; affective sounds were used because robot speech influences proxemic preferences and would have introduced a confound in the experiment [29].

4 Experimental Design

With the described experimental setup, we performed an experiment to investigate user perceptions of robot performance attenuated by distance and its effect on proxemic preferences.

4.1 Experimental Procedure

Participants (described in Section 4.4) were greeted at the door entering the private experimental space, and were informed of and agreed to the nature of the experiment and their rights as a participant, which included a statement that the experiment could be halted at any time.

Participants were then instructed to stand with their toes touching a line on the floor, and were asked to remain there for the duration of the experiment [Figure 2]. The experimenter then provided instructions about the task the participant would be performing.

Participants were introduced to the robot, and were informed that all of its actions were completely autonomous. Participants were told that the robot would be moving along a straight line throughout the duration of the experiment; a brief demonstration of robot motion was provided, in which the robot autonomously moved back and forth between distances of 3.0 meters and 4.5 meters from the participant, allowing them to familiarize themselves with the robot motion. Participants were told that they would be asked about some of their preferences regarding the robot’s location throughout the experiment.

Participants were then informed that they would be instructing the robot to look at any one of four boxes (of their choosing) located in the room [Figure 2], and that they could use speech (in English) and pointing gestures. A vocabulary for robot instructions was provided: for speech, participants were told they could say the words ”look at” followed by the name of the shape or color of each box (e.g., ”triangle”, ”circle”, ”blue”, ”purple”, etc.); for pointing gestures, participants were asked to use their left arm to point to boxes located on their left, and their right arm to point to boxes on their right. This vocabulary was provided to minimize any perceptions the person might have that the robot simply did not understand the words or gestures that they used; thus, the use of the vocabulary attempted to maximize the perception that any failures of the robot were due to other factors.

Participants were told that they would repeat this instruction procedure to the robot many times, and that the robot would indicate whether or not it understood their instructions each time using the head movements and audio cues described in Section 3.2.

Participants had an opportunity to ask the experimenter any clarifying questions. Once participant understanding was verified, we proceeded with the experiment.

4.1.1 Pre-interaction Proxemic Measures (*pre*)[7]

The robot autonomously moved to the “far home” location [Figure 2]. Participants were told that the robot would be approaching them, and to say out loud the word “stop” when the robot reached the ideal location at which the participant would have a *face-to-face conversation*[8] with the robot. This pre-interaction proxemic preference from the “far home” location is denoted as pre_{far}.

When the participant was ready, the experimenter pressed a PS3 button to start the robot moving. When the participant said “stop”, the experimenter pressed another button to halt robot movement. The experimenter pressed another button to record the distance between the robot and the participant, as measured by the on-board laser.

Once the pre_{far} distance was recorded, the experimenter pressed another button, and the robot autonomously moved to the “near home” location [Figure 2]; the participant was informed that the robot would be approaching to this location and would stop on its own. The process was repeated with the robot backing away from the participant, and the participant saying “stop” when it reached the ideal location for conversation. This pre-interaction proxemic preference from the “near home” location is denoted as pre_{near}.

[7] Measures are provided inline with the experimental procedure to provide an order of events as they occurred in the experiment.

[8] Related work in human-robot proxemics asks the participant about locations at which they feel *comfortable* [24], yielding proxemic preferences very near to the participant. Our general interest is in face-to-face human-robot conversational interaction, with proxemic preference farther from the participant [16, 26, 27], hence the choice of wording.

From pre_{far} and pre_{near}, we calculated and recorded the average pre-interaction proxemic preference, denoted as pre[9].

4.1.2 *Interaction Scenario*

After determining pre-interaction proxemic preferences, the robot returned to the "far home" location. The experimenter then repeated to participants the instructions about the task they would be performing with the robot. When participants verified that they understood the task and indicated that they were ready, the experimenter pressed a button to proceed with the task.

The robot autonomously visited ten pre-determined locations [Figure 2]. At each location, the robot responded to instructions from the participant to look at one of four boxes located in the room [Figure 2]. Five instruction-response interactions were performed at each location, after which the robot moved to the next location along its path; thus, each participant experienced a total of 50 instruction-responses interactions. Robot goal locations were in 0.5-meter intervals inclusively between the "near home" location (0.25 meters) and "far home" location (4.75 meters) along a straight-line path in front of the participant [Figure 2]. Locations were visited in a sequential order; for half of the participants, the robot approached from the "far home" location (i.e., farthest-to-nearest order), and, for the other half of participants, the robot backed away from "near home" location (i.e., nearest-to-farthest order); this was done to reduce any ordering effects [19].

To controllably simulate social signal attenuation at each location, robot performance was artificially manipulated as a function of the distance to the participant (described in Section 4.2). After each instruction provided by the participant, the experimenter provided to the robot (via a remote control interface) the ground-truth of the instruction; the robot then determined whether or not it would have understood the instruction based on a prediction from a performance vs. distance curve (specified by the assigned experimental condition described in Section 4.2), and provided either an *affirmative response* or a *negative response* to the participant indicating its successful or failed understanding of the instruction, respectively.

The entire interaction scenario lasted 10-15 minutes.

4.1.3 *Post-interaction Proxemic Measures ($post$)*

After the robot visited each of the ten locations, it autonomously returned to the "far home" location. The experimenter then repeated the procedure for determining proxemic preferences described in Section 4.1.1. This process generated post-performance proxemic preferences from the "far home" and "near home" locations, as well as their average, denoted $post_{far}$, $post_{near}$, and $post$[10], respectively.

4.1.4 *Perceived Peak Location Measures ($perc$)*

Finally, after collecting post-interaction proxemic preferences, the experimenter repeated the procedure described in Section 4.1.1 to determine participant perceptions of the location of peak performance. This process generated perceived peak performance locations from the "far home" and "near home" locations, as well as their average, denoted $perc_{far}$, $perc_{near}$, and $perc$[11], respectively.

[9] Post-hoc analysis revealed no statistically significant difference between pre_{far} and pre_{near} measurements, hence why we rely on pre.

[10] Post-hoc analysis revealed no statistically significant difference between $post_{far}$ and $post_{near}$ measurements, hence why we rely on $post$.

[11] Post-hoc analysis revealed no statistically significant difference between $perc_{far}$ and $perc_{near}$ measurements, hence why we rely on $perc$.

4.2 Experimental Conditions

We considered two performance vs. distance conditions; 1) a **"uniform performance" condition**, and 2) an **"attenuated performance" condition**. Overall robot performance for each condition was held at a constant 40%[12]—that is, for each participant, the robot provided 20 affirmative responses and 30 negative responses distributed across 50 instructions. The way in which these responses were distributed across locations varied between conditions.

In the **uniform performance condition**, robot performance was the same (40%) across across all locations [Figures 3 and 4]. Thus, at each of the ten locations visited, the robot provided two affirmative and three negative responses, respectively. This condition served as a baseline of participant proxemic preferences within the task.

In the **attenuated performance condition**, robot performance varied with distance proportional to a Gaussian distribution centered a location of "peak performance" ($M = peak$, $SD = 1.0$) [Figures 3 and 4]. Due to differences in pre-interaction proxemic preferences, we could not select a single value for $peak$ that provided a similar experience between participants without introducing other confounding factors (e.g., the $peak$ not being at a location that the robot visited or distances beyond the "home" locations). To alleviate this, we opted to select multiple peak performance locations, exploring the space of human responses to robot performance differences at a variety of distances. We selected the eight locations non-inclusively between the "near home" and "far home" locations as the peak performance locations [Figure 2]; the "near home" and "far home" locations were not included in the set of peaks to ensure that participants were always exposed to an actual *peak* in performance, rather than just a *trend*. Peak performance locations were varied between participants.

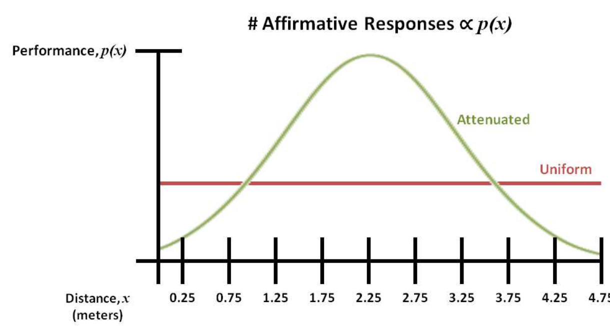


Figure 3. The performance curves of the **uniform** and **attenuated** conditions. In this example, $peak = 2.25$ (in meters), so the attenuated performance curve parameters is $M = peak = 2.25$, $SD = 1.0$. The number of affirmative responses at a distance, x, from the user is proportional to $p(x)$, the evaluation of the performance curve at x.

The distribution of affirmative responses for all conditions is presented in Figure 4. The number of affirmative responses was normalized to 20 (40%) to ensure a consistent user experience of overall robot performance across all conditions. In the **attenuated performance condition**, the number of affirmative responses at $peak$ was always the 5 (i.e., perfect performance), and the number of affirmative responses at other locations were always less than that of the peak to ensure that participants were exposed to an actual peak. At each location, the order in which the five responses were provided was random.

[12] This value was selected because it is an average performance rate predicted by our results in [16] for typical human-robot proxemic preferences.

			# Affirmative Responses vs. Distance									
	Distance, x (meters) →		0.25	0.75	1.25	1.75	2.25	2.75	3.25	3.75	4.25	4.75
Performance Condition	***Uniform*** →		2	2	2	2	2	2	2	2	2	2
	Attenuated *peak* → (meters)	***0.75***	4	***5***	4	3	1	1	1	1	0	0
		1.25	3	4	***5***	4	3	1	0	0	0	0
		1.75	1	3	4	***5***	4	3	0	0	0	0
		2.25	0	1	3	4	***5***	4	3	0	0	0
		2.75	0	0	0	3	4	***5***	4	3	1	0
		3.25	0	0	0	0	3	4	***5***	4	3	1
		3.75	0	0	0	0	1	3	4	***5***	4	3
		4.25	0	0	1	1	1	1	3	4	***5***	4

Figure 4. The distribution of affirmative responses provided by the robot across conditions. Manipulated values are highlighted in ***bold italics***.

4.3 Experimental Hypotheses

Within these conditions, we had three central hypotheses:

H1: In the **uniform performance condition**, there will be no significant change in participant proxemic preferences.

H2: In the **attenuated performance conditions**, participants will be able to identify a relationship between robot performance and human-robot proxemics.

H3: In the **attenuated performance conditions**, participants will adapt their proxemic preferences to improve robot performance.

4.4 Participants

We recruited 100 participants (50 male, 50 female) from our university campus community. Participant race was diverse (67 white/Caucasian, 26 Asian, 3 black/African-American, 3 Latino/Latina, and 1 mixed-race). All participants reported proficiency in English and had lived in the United States for at least two years (i.e., acclimated to U.S. culture). Average age (in years) of participants was 22.26 ($SD = 4.31$), ranging from 18 to 39. Based on a seven-point scale, participants reported moderate familiarity with technology ($M = 3.98, SD = 0.85$). Average participant height (in meters) was 1.74 ($SD = 0.10$), ranging from 1.52 to 1.93. Related work reports how human-robot proxemics is influenced by gender and technology familiarity [24], culture [3], and height [7, 21].

The 100 participants were randomly assigned to a performance condition, with $N = 20$ in the **uniform performance condition** and $N = 80$ in the **attenuated performance condition**. In the **attenuated performance condition**, the 80 participants were randomly assigned one of the eight peak performance locations (described in Section 4.2) with $N = 10$ for each $peak$. Neither the participant nor the experimenter was aware of the condition assigned.

5 Results and Discussion

We analyzed data collected in our experiment to test our three hypotheses (described in Section 4.3), and evaluated their implications for autonomous social robots and human-robot proxemics.

To provide a baseline of our robot for comparison in general human-robot proxemics, we consolidated and analyzed pre-interaction proxemic preferences (pre) across all conditions ($N = 100$), as the data had not been influenced by robot performance. The participant pre-interaction proxemic preference (in meters) was determined to be 1.14 ($SD = 0.49$) for our robot system, which is consistent with [18] and our previous work [16], but twice as far away as related work has reported for robots of a similar form factor [28, 24].

5.1 H1: Pre- vs. Post-interaction Locations

To test **H1**, we compared average pre-interaction proxemic preferences (pre) to average post-interaction proxemic preferences ($post$) of participants in the **uniform performance condition**.

A paired t-test revealed a statistically significant change in participant proxemic preferences between pre ($M = 1.12, SD = 0.51$) and $post$ ($M = 1.39, SD = 0.63$); $t(38) = 1.49, p = 0.02$. Thus, our hypothesis **H1** is rejected.

The rejection of this hypothesis does not imply a failure of the experimental procedure, but, rather, provides important insights that must be considered for subsequent analyses (and for related work in proxemics). This result suggests that there might be something about the context of the interaction scenario itself that influenced participant proxemic preferences. To address any influence the interaction scenario might have on subsequent analyses, we define a *contextual offset*, θ, as the average difference in participant post-interaction and pre-interaction proxemic preferences ($M = 0.27, SD = 0.48$); this θ value will be subtracted from ($post - pre$) values in Section 5.3 to normalize for the interaction context.

5.2 H2: Perceived vs. Actual Peak Locations

To test **H2**, we compared participant perceived locations of peak performance ($perc$) to actual locations of peak performance ($peak$) in the **attenuated performance conditions** [Figure 5].

Steven's Power Law, ax^b, has previously been used to model human distance estimation as a function of actual distance [19], and is generally well representative of human-perceived vs. actual stimuli [23]. However, existing Power Laws relevant to our work only seem to pertain to distances of 3–23 meters, which are beyond the range of the natural face-to-face communication with which we are concerned. Thus, our goal here is to model our own experimental data to establish a Power Law for $perc$ vs. $peak$ at locations more relevant to HRI (0.75–4.25 meters), which we can then evaluate to test **H2**.

Immediate observations of our data suggested that the data appear to be heteroscedastic [Figure 5]—in this case, the variance seems to increase with distance from the participant, which means we should not use traditional statistical tests. The Breusch-Pagan test for non-constant variance (NCV) confirmed this intuition; $\chi^2(1, N = 100) = 15.79, p < 0.001$. A commonly used and accepted approach to alleviate our heteroscedasticity is to transform the $perc$ and $peak$ data to a log-log scale. While not applicable to all datasets, this approach served as an adequate approximation for our purposes [Figure 6]; it also enabled us to perform a regression analysis to determine parameter values for the Power Law coefficient and exponent, $a = 1.3224$ and $b = 0.5132$, respectively. With these parameters, we identified that $peak$ was a strongly correlated and very significant predictor of $perc$; $R^2 = 0.4951, F(1, 78) = 76.48, p < 0.001$. Thus, our hypothesis **H2** is supported.

This result suggests that people are able to identify a relationship between robot performance and human-robot proxemics, but they will predictably underestimate the distance, x, to the location of peak performance based on the Power Law equation $1.3224x^{0.5132}$. While human estimation of the location of peak performance is suboptimal, it is possible that repeated exposure to the robot over multiple sessions might yield more accurate results. This follow-up hypothesis will be formally tested in a planned longitudinal study in future work (described in Section 6).

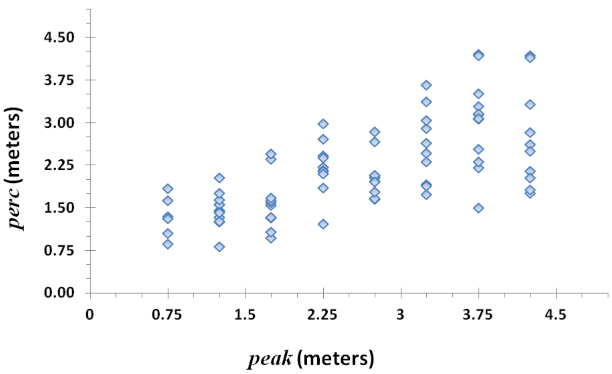


Figure 5. Participant perceived location of robot peak performance ($perc$) vs. actual location of robot peak performance ($peak$). Note the heteroscedasticity of the data, which prevents us from performing traditional statistical analyses without first transforming the data (shown in Figure 6).

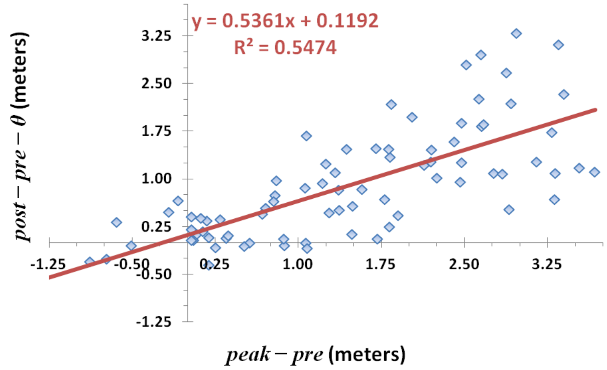


Figure 7. Changes in participant pre-/post-interaction proxemic preferences (pre and $post$, respectively; θ is the *contextual offset* defined in Section 5.1) vs. distance from participant pre-interaction proxemic preference (pre) to the actual location of robot peak performance ($peak$).

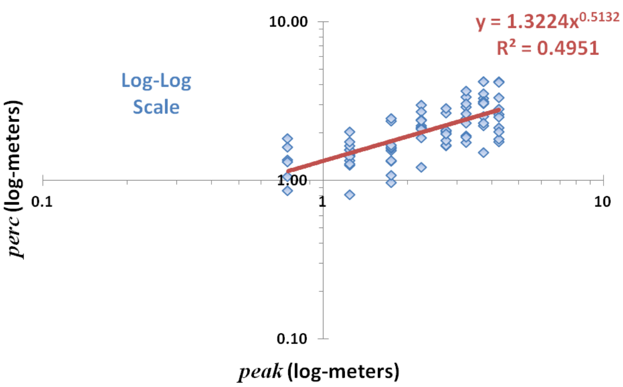


Figure 6. Participant perceived location of robot peak performance ($perc$) vs. actual location of robot peak performance ($peak$) on a log-log scale, reducing the effects of heteroscedasticity and allowing us to perform regression to determine parameters of the Power Law, ax^b.

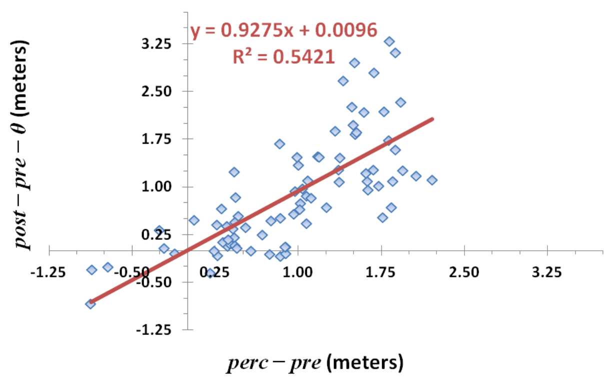


Figure 8. Changes in participant pre-/post-interaction proxemic preferences (pre and $post$, respectively; θ is the *contextual offset* defined in Section 5.1) vs. distance from participant pre-interaction proxemic preference (pre) to the perceived location of robot peak performance ($perc$).

5.3 H3: Preferences vs. Peak Locations

To test **H3**, we compared changes in participant pre-/post-interaction proxemic preferences ($post - pre - \theta$) to the distance from the participant pre-interaction proxemic preference to either a) the actual location of robot peak performance ($peak - pre$) [Figure 7], or b) the perceived location of robot peak performance ($perc - pre$) [Figure 8], both in the **attenuated performance conditions**.

Data for ($post - pre - \theta$) vs. both ($peak - pre$) and ($perc - pre$) were heteroscedastic, as indicated by Breusch-Pagan NCV tests: $\chi^2(1, N = 100) = 18.81, p < 0.001$; and $\chi^2(1, N = 100) = 13.55, p < 0.001$; respectively. This is intuitive, as the data for perceived ($perc$) vs. actual ($peak$) locations of peak performance were also heteroscedastic [Figure 5]. The log-transformation approach that we used in Section 5.2 did not perform well in modeling these data; thus, we needed to use an alternative approach. We opted to utilize a Generalized Linear Model [20] because it allowed us to model the variance of each measurement separately as a function of predicted values and, thus, perform appropriate statistical tests for significance.

We first modeled changes in participant proxemic preferences ($post - pre - \theta$) vs. distance from pre-interaction proxemic preference to the actual location of peak performance ($peak - pre$). In the ideal situation (for the robot), these match one-to-one—in other words, the participant meets the needs of the robot entirely by changing proxemic preferences to be centered at the peak of robot performance. Unfortunately for the robot, this was not the case. We detected a strongly correlated and statistically significant relationship between participant proxemic preference change and distance from pre-interaction preference to the peak location ($R^2 = 0.5474, \beta = 0.5361, t(98) = 9.71, p < 0.001$), but participant preference change only got the robot approximately halfway ($\beta = 0.5361$) to its location of peak performance [Figure 7]. Why is this?

Recall that results reported in Section 5.2 suggested that, while people do perceive a relationship between robot performance and distance, their ability to accurately identify the location of robot peak performance diminishes based on the distance to it as governed by a Power Law. Were participants *trying* to maximize robot performance, but simply adapting their preferences to a suboptimal location?

We investigated this question by considering changes in participant proxemic preferences ($post - pre - \theta$) vs. distance from pre-interaction proxemic preference to the perceived location of peak performance ($perc - pre$). If the participant was adapting their proxemic preferences to accommodate the needs of the robot, then these should match one-to-one. A Generalized Linear Model was fit to these data, and yielded a strongly correlated and statistically significant relationship between changes in proxemic preferences and perceptions of robot performance ($R^2 = 0.5421, \beta = 0.9275, t(98) = 9.61, p < 0.001$) [Figure 8]. Thus, our hypothesis **H3** is supported.

The near one-to-one relationship ($\beta = 0.9275$) between post-interaction proxemic preferences and participant perceptions of robot peak performance is compelling, suggesting that participants adapted their proxemic preferences almost entirely to improve robot performance in the interaction.

5.4 Discussion

These results have significant implications for the design of social robots and autonomous robot proxemic control systems, specifically, in that people's proxemic preferences will likely change as the user interacts with and comes to understand the needs of the robot.

As illustrated in our previous work [16], the locations of on-board sensors for social signal recognition (e.g., microphones and cameras), as well as the automated speech and gesture recognition software used, can have significant impacts on the performance of the robot in autonomous face-to-face social interactions. As our now-reported results suggest that people will adapt their behavior in an effort to improve robot performance, it is anticipated that human-robot proxemics will vary between robot platforms with different hardware and software configurations based on factors that are 1) not specific to the user (unlike culture [3], or gender, personality, or familiarity with technology [24]), 2) not observable to the user (unlike height [7, 21], amount of eye contact [24, 18], or vocal parameters [29]), or 3) not observable to the robot developer. User understanding of the relationship between robot performance and human-robot proxemics is a latent factor that only develops through repeated interactions with the robot (perhaps expedited by the robot communicating its predicted error); fortunately, our results indicate that user understanding will develop in a predictable way. Thus, it is recommended that social robot developers consider and perhaps model robot performance as a function of conditions that might occur in dynamic proxemic interactions with human users to better predict and accommodate how the people will actually use the technology. This dynamic relationship, in turn, will enable more rich autonomy for social robots by improving the performance of their own automated recognition systems.

If developers adopt models of robot performance as a factor contributing to human-robot proxemics, then it follows that proxemic control systems might be designed to expedite the process of autonomously positioning the robot at an optimal distance from the user to maximize robot performance while still accommodating the initial personal space preferences of the user. This was the focus of our previous work [16], which treated proxemics as an optimization problem that considers the production and perception of social signals (speech and gesture) as a function of distance and orientation. Recall that an objective of the now-reported work was to address questions regarding whether or not users would accept a robot that positions itself in locations that might differ from their initial proxemic preferences. The results in this work (specifically, in Section 5.3) support the notion that user proxemic preferences will change through interactions with the robot as its performance is observed, and that the new user proxemic preference will be at the *perceived* location of robot peak performance. An extension of this result is that, through repeated interactions, user proxemic preferences will further adapt and eventually converge to the *actual* location of robot peak performance, a hypothesis that we will investigate in future work.

6 Future Work

Our experimental conditions (described in Section 4.2) were specifically selected to strongly expose a relationship (if one existed) between human proxemic preferences and robot performance—the robot achieved perfect success rates (100%) at "peak" locations and perfect failure rates (0%) at other locations, and these success/failure rates were distributed proportional to a Gaussian distribution with constant variance. Now that we have identified that a relationship exists, our next steps will examine how the relationship changes over time or with other related factors. A longitudinal study over multiple sessions will be conducted to determine if changes in preferences persist from one interaction to the next, and if user proxemic preferences will continue to adapt and eventually converge to locations of robot peak performance through repeated interactions. Other future work will follow the same experimental procedure described in Section 4.1, but will adjust the **attenuated performance condition** (described in Section 4.2) to consider how the relationship changes with 1) distributions of lower or higher variance, 2) lower maximum performance or higher minimum performance, 3) more realistic non-Gaussian distributions, and 4) the interactions between distributions of actual multimodal recognition systems [16].

This perspective opens up a whole new theoretical design space of human-robot proxemic behavior. The general question is, "How will people adapt their proxemic preferences in any given *performance field*?", in which performance varies with a variety of factors, such as distance, orientation, and environmental interference. The follow-up question then asks, "How can the robot expedite the process of establishing an appropriate human-robot proxemic configuration within the performance field without causing user discomfort?" This will be a focus of future work, and will extend our prior work on modeling human-robot proxemics to improve robot proxemic controllers [16].

7 Summary and Conclusions

An objective of autonomous socially assistive robots is to meet the needs and preferences of a human user [4]. However, this can sometimes be at the expense of the robot's own ability to understand social signals produced by the user. In particular, human proxemic preferences with respect to a robot can have significant impacts on the performance rates of its automated speech and gesture recognition systems [16]. This means that, for a successful interaction, the robot has needs too—and these needs might not be consistent with and might require changes in the proxemic preferences of the human user.

In this work, we investigated how user proxemic preferences changed to improve the robot's understanding of human social signals (described in Section 4). We performed an experiment in which a robot's performance was artificially varied, either *uniformly* or *attenuated* across distance. Participants ($N = 100$) instructed a robot using speech and pointing gestures, and provided their proxemic preferences before and after the interaction.

We report two major findings. First, people predictably underestimate the distance to the location of robot peak performance; the relationship between participant perceived and actual distance to the location of peak performance is represented well by a Power Law (described in Section 5.2). Second, people adjust their proxemic preferences to be near the *perceived* location of maximum robot understanding (described in Section 5.3). This work offers insights into the dynamic nature of human-robot proxemics, and has significant implications for the design of social robots and robust autonomous robot proxemic control systems (described in Section 5.4).

Traditionally, we focus on our attention on ensuring the robot is meeting the needs of the user with little regard to the impact it might have on the robot itself; it is often an afterthought, or something that we, as robot developers, have to "fix" with our systems. While robot developers will continue to improve upon our autonomous systems, our results suggest that even novice users are willing to adapt their behaviors in an effort to help the robot better understand and perform its tasks. Automated recognition systems are not and will likely never be perfect, but this is no reason to delay the development, deployment, and benefits of social and socially assistive robot technologies. Robots have needs too, and human users will attempt to meet them.

ACKNOWLEDGEMENTS

This work is supported in part by an NSF Graduate Research Fellowship, the NSF National Robotics Initiative (IIS-1208500), and an NSF CNS-0709296 grant.

We thank Aditya Bhatter, Lizhi Fan, Jonathan Frey, Akash Metawala, Kedar Prabhu, and Cherrie Wang for their assistance in recruiting participants and conducting the experiment.

REFERENCES

[1] C. Breazeal, 'Social interactions in hri: The robot view', *IEEE Transactions on Man, Cybernetics and Systems*, **34**(2), 181–186, (2003).
[2] C. Breazeal, *Designing Sociable Robots*, MIT Press, Cambridge, Massachusetts, 2004.
[3] G. Eresha, M. Haring, B. Endrass, E. Andre, and M. Obaid, 'Investigating the influence of culture on proxemic behaviors for humanoid robots', in *22nd IEEE International Symposium on Robot and Human Interactive Communication*, RO-MAN 2013, pp. 430–435, (2013).
[4] D.J. Feil-Seifer and M.J. Matarić, 'Defining socially assistive robotics', in *International Conference on Rehabilitation Robotics*, ICRR'05, pp. 465–468, Chicago, Illinois, (2005).
[5] E. T. Hall, *The Silent Language*, Doubleday Company, New York, New York, 1959.
[6] E.T Hall, 'A system for notation of proxemic behavior', *American Anthropologist*, **65**, 1003–1026, (1963).
[7] Y. Hiroi and A. Ito, *Influence of the Size Factor of a Mobile Robot Moving Toward a Human on Subjective Acceptable Distance*.
[8] P.H. Kahn, *Technological Nature: Adaptation and the Future of Human Life*, MIT Press, Cambridge, Massachusetts, 2011.
[9] D. Koller and N. Friedman, *Probabilistic Graphical Models*, MIT Press, Cambridge, Massachusetts, 2009.
[10] R. Mead, 'Space, speech, and gesture in human-robot interaction', in *Doctoral Consortium of International Conference on Multimodal Interaction*, ICMI'12, pp. 333–336, Santa Monica, California, (2012).
[11] R. Mead, A. Atrash, and M. J. Matarić, 'Proxemic feature recognition for interactive robots: Automating metrics from the social sciences', in *International Conference on Social Robotics*, pp. 52–61, Amsterdam, Netherlands, (2011).
[12] R. Mead, A. Atrash, and M. J. Matarić, 'Recognition of spatial dynamics for predicting social interaction', in *6th ACM/IEEE International Conference on Human-Robot Interaction*, pp. 201–202, Lausanne, Switzerland, (2011).
[13] R. Mead, A. Atrash, and M. J. Matarić, 'Representations of proxemic behavior for human-machine interaction', in *NordiCHI 2012 Workshop on Proxemics in Human-Computer Interaction*, NordiCHI'12, Copenhagen, Denmark, (2012).
[14] R. Mead, A. Atrash, and M. J. Matarić, 'Automated proxemic feature extraction and behavior recognition: Applications in human-robot interaction', *International Journal of Social Robotics*, **5**(3), 367–378, (2013).
[15] R. Mead and M. J. Matarić, 'A probabilistic framework for autonomous proxemic control in situated and mobile human-robot interaction', in *7th ACM/IEEE International Conference on Human-Robot Interaction*, HRI'12, pp. 193–194, Boston, Massachusetts, (2012).
[16] R. Mead and M. J. Matarić, 'Perceptual models of human-robot proxemics', in *14th International Symposium on Experimental Robotics*, ISER'14, p. to appear, Marrakech/Essaouira, Morocco, (2014).
[17] R. Mead, E. Wade, P. Johnson, A. St. Clair, S. Chen, and M. J. Mataric., 'An architecture for rehabilitation task practice in socially assistive human-robot interaction', in *Robot and Human Interactive Communication*, pp. 404–409, (2010).
[18] J. Mumm and B. Mutlu, 'Human-robot proxemics: Physical and psychological distancing in human-robot interaction', in *6th ACM/IEEE International Conference on Human-Robot Interaction*, HRI-2011, pp. 331–338, Lausanne, (2011).
[19] A Murata, 'Basic characteristics of human's distance estimation', in *1999 IEEE International Conference on Systems, Man, and Cybernetics*, volume 2 of *SMC'99*, pp. 38–43, (1999).
[20] J. Nelder and R. Wedderburn, 'Generalized linear models', *Journal of the Royal Statistical Society*, **135**(3), 370–384, (1972).
[21] I. Rae, L. Takayama, and B. Mutlu, 'The influence of height in robot-mediated communication', in *8th ACM/IEEE International Conference on Human-Robot Interaction*, HRI-2013, pp. 1–8, Tokyo, Japan, (2005).
[22] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, 'How to approach humans?: Strategies for social robots to initiate interaction', in *HRI*, pp. 109–116, (2009).
[23] S.S. Stevens, 'On the psychological law', *Psychological Review*, **64**, 153–181, (2007).
[24] L. Takayama and C. Pantofaru, 'Influences on proxemic behaviors in human-robot interaction', in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, IROS'09, pp. 5495–5502, (2009).
[25] A. Tapus, M.J. Matarić, and B. Scassellati, 'The grand challenges in socially assistive robotics', *IEEE Robotics and Automation Magazine*, **14**(1), 35–42, (2007).
[26] Elena Torta, Raymond H. Cuijpers, and James F. Juola, 'Design of a parametric model of personal space for robotic social navigation', *International Journal of Social Robotics*, **5**(3), 357–365, (2013).
[27] Elena Torta, Raymond H. Cuijpers, James F. Juola, and David van der Pol, 'Design of robust robotic proxemic behaviour', in *Proceedings of the Third International Conference on Social Robotics*, ICSR'11, pp. 21–30, (2011).
[28] M.L. Walters, K. Dautenhahn, R.T. Boekhorst, K.L. Koay, D.S. Syrdal, and C.L.: Nehaniv, 'An empirical framework for human-robot proxemics', in *New Frontiers in Human-Robot Interaction*, pp. 144–149, Edinburgh, (2009).
[29] M.L. Walters, D.S. Syrdal, K.L. Koay, K. Dautenhahn, and R. te Boekhorst, 'Human approach distances to a mechanical-looking robot with different robot voice styles', in *The 17th IEEE International Symposium on Robot and Human Interactive Communication*, RO-MAN 2008, pp. 707–712, (2008).

4th International Symposium on New Frontiers in Human-Robot Interaction
21–22 April 2015, Canterbury, UK
M. Salem, A. Weiss, P. Baxter, K. Dautenhahn (Eds.)

A new biomimetic approach towards educational robotics: the Distributed Adaptive Control of a Synthetic Tutor Assistant

Vasiliki Vouloutsi[1], Maria Blancas Munoz[1], Klaudia Grechuta[1], Stephane Lallee[1,2], Armin Duff[1], Jordi-Ysard Puigbo Llobet[1] and Paul F.M.J. Verschure[1,3]

Abstract. Many fields can profit from the introduction of robots, including that of education. In this paper, our main focus is the advancement of the Synthetic Tutor Assistant (STA), a robot that will act as a peer for knowledge transfer. We propose a theory of a tutoring robotic application that is based on the Distributed Adaptive Control (DAC) theory: a layered architecture that serves as the framework of the proposed application. We describe the main components of the STA and we evaluate the implementation within an educational scenario.

1 INTRODUCTION

Robots are now able to interact with humans in various conditions and situations. Lately, there has been an increased attempt to develop socially interactive robots, that is, robots with the ability to display social characteristics: use natural communicative cues (such as gestures or gaze), express emotional states or even establish social relationships, all of which are important when a peer-to-peer interaction takes place [20]. In fact, given the current technological advancements, we are now able to develop robotic systems that are able to deal with both physical and social environments. One of the greatest challenges in the design of social robots is to correctly identify all those various factors that affect social interaction and act in accordance [43]. Indeed, different studies have shown that the complexity in the behavior of robots affect how humans interact with robots and perceive them [30, 55, 7, 52].

There are many fields that can profit from the introduction of robots [13], they include health care [9], entertainment [18], social partners [8] or education [21, 41]. Here we focus on the latter, by advancing the notion of the Synthetic Tutor Assistant (STA) (see section 3) which is pursued in the European project entitled Expressive Agents for Symbiotic Education and Learning (EASEL). In this perspective, the robot STA will not act as the teacher, but rather as a peer of the learner to assist in knowledge acquisition. It has been shown that robots can both influence the performance of the learner [41] and their motivation to learn [29]. One of the main advantages of employing a robotic tutor is that it can provide assistance at the level of individual learners, given that the robot can have the ability to learn and adapt based on previous interactions.

[1] Laboratory of Synthetic Perceptive, Emotive and Cognitive Systems - SPECS, Universitat Pompeu Fabra
[2] Institute For Infocomm Research, A*STAR, Singapore
[3] Center of Autonomous Systems and Neurorobotics - NRAS, Catalan Institute of Advanced Studies - ICREA, email: paul.verschure@upf.edu

Through education, people acquire knowledge, develop skills and capabilities and consequently form values and habits. Although there exist several educational approaches that could be considered, here, we will focus on Constructivism [35]. Constructivism proposes an educational approach based on collaboration, learning through making, and technology-enhanced environments. Such approach aims at constructing social interaction between the participant and the STA as it implies a common goal for both learners-players [45].

We consider tutoring as the structured process in which knowledge and skills are transferred to an autonomous learner through a guided process based on the individual traits of the learner. Here we present an approach where both the user model and the STA are based on a neuroscientifically grounded cognitive architecture called Distributed Adaptive Control (DAC) [51, 47]. On one hand, DAC serves as the theory which defines the tutoring scenario: it allows us to derive a set of key principles that are general for all learning processes. On the other hand, it is the core for the implementation of the control architecture of the STA, the robotic application. Following the layered DAC architecture, we propose the STA that will deploy tutoring strategies of increasing levels of complexity depending on the performance and capabilities of the learner. The DAC theory serves as both the basis for the tutoring robotic application, user model as well as for the implementation of the STA. Such design guarantees a tight interplay between the robotic application, the user and their interaction.

The present study is organized as follows: first, we present the background theory of the tutoring robotic application, the DAC theory, and we describe the tutoring model applied. Furthermore, we introduce the key implementation features of the STA based on DAC. To assess the first implementation of our system, we devised a pilot study where the STA performs the role of a peer-teacher in an educational scenario. The proposed scenario consists of a pairing game where participants have to match an object to its corresponding category. The setup was tested with both children and adults. The game had three levels of increased difficulty. Questionnaires distributed after every interaction to the players were used to assess the STA's ability to transfer knowledge.

2 DAC COGNITIVE ARCHITECTURE AND LEARNING

To provide a model of perception, cognition and action for our system, we have implemented the DAC architecture. [51, 47]. DAC is a theory of mind and brain, and its implementation serves as a real-time neuronal model for perception, cognition and action (for a review see [49]). DAC will serve both as the basis for the tutoring model as well as the core of the implementation of the STA.

2.1 Distributed Adaptive Control (DAC)

Providing a real-time model for perception, cognition and action, DAC has been formulated in the context of classical and operant conditioning: learning paradigms for sensory-sensory, multi-scale sensorimotor learning and planning underlying any form of learning. According to DAC, the brain is a layered control architecture that is subdivided into functional segments sub-serving the processing of the states of the world, the self, interaction through action [48], and it is dominated by parallel and distributed control loops.

DAC proposes that in order to act upon the environment (or to realize the *How?* of survival) the brain has to answer four fundamental questions, continuously and in real-time: *Why, What, Where and When*, forming the H4W problem [50, 49]. However, in a world filled with agents, the H4W problem does not seem enough to ensure survival; an additional key question needs to be answered: *Who?*, which shifts the H4W towards a more complex problem, H5W [46, 39].

To answer the H5W problem, the DAC architecture comprises of four layers: Somatic, Reactive, Adaptive and Contextual, intersected by three columns: states of the world (exosensing), states of self (endosensing) and their interface in action (Figure 1). The Somatic Layer represents the body itself and the information acquired from sensations, needs and actions. The Reactive Layer comprises fast, predefined sensorimotor loops (reflexes) that are triggered by low complexity perceptions and are coupled to specific affective states of the agent. It supports the basic functionality of the Somatic Layer in terms of reflexive behavior and constitutes the main behavioral system based on the organism's physical needs. Behavior emerges from the satisfaction of homeostatic needs, which are also regulated by an integrative allostatic loop that sets the priorities and hierarchies of all the competitive homeostatic systems. Thus, behavior serves the reduction of needs [25] controlled by the allostatic controller [42].

The Adaptive Layer extends the sensorimotor loops of the Reactive Layer with acquired sensor and action states, allowing the agent to escape the predefined reflexes and employs mechanisms to deal with unpredictability through learning [14]. The Contextual Layer uses the state-space acquired by the Adaptive Layer to generate goal oriented behavioral plans and policies. This layer includes mechanisms for short, long-term and working memory, formating sequential representations of states of the environment and actions generated by the agent or its acquired sensorimotor contingencies. The DAC architecture has been validated through robotic implementations[19, 42], expanded to capture social interactions with robots [52, 39] as well as providing novel approaches towards rehabilitation [47]. Here, the implementation of DAC serves two main purposes. On the one hand, it acts as the grounding theory for the pedagogical model: it allows us to derive and deduce a set of key principles that are general for all learning processes. On the other hand, DAC is the core for the implementation of the STA.

2.2 Phases of learning

Based on the formal description of learning from the DAC architecture which has been shown to be Bayes optimal [48], we will focus on two main principles as it has a dual role within EASEL. On one hand, DAC is the core for the implementation of the Synthetic Tutor

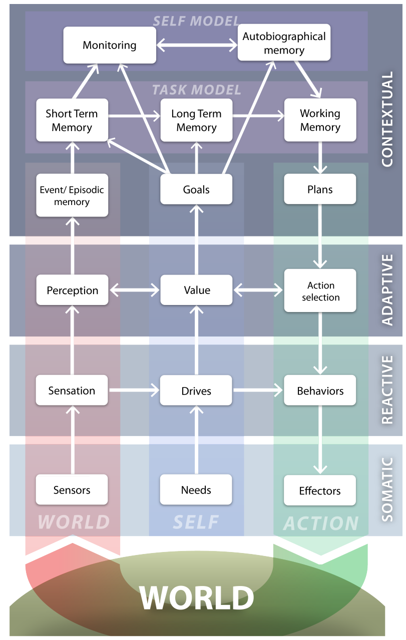


Figure 1. The DAC architecture and its four layers (somatic, reactive adaptive and contextual). Across the layers we can distinguish three functional columns of organization: world (exosensing), self (endosensing) and action (the interface to the world through action). The arrows show the flow of information. Image adapted from [49].

Assistant (STA). On the other hand, following the layered architecture, the STA deploys pedagogical strategies of increasing levels of complexity.

First, DAC predicts that learning is bootstrapped and organized along a hierarchy of complexity: the Reactive Layer allows for exploring the world and gaining experiences, based on which the Adaptive Layer learns the states of the world and their associations; only after these states are well consolidated, the Contextual Layer can extract consistent rules and regularities. We believe that the same hierarchy is applicable in the pedagogical context. Secondly, DAC predicts that in order to learn and consolidate new material, the learner undergoes a sequence of learning phases: resistance, confusion and resolution. *Resistance* is a mechanism that results from defending one's own (in)competence level against discrepancies encountered in sensor data. In DAC these discrepancies regulate both perceptual learning and the engagement of sequence memory. Consistent perceptual and behavioral errors lead to the second phase, namely *confusion*, the necessity to resolve the problem and learn through readapting. *Confusion* modulates learning as to facilitate the discovery and generation of new states to be assessed on their validity. In other words, to assist in performing abduction. Finally, *resolution* is the very process of stabilizing new knowledge that resolves the earlier encountered discrepancies and errors. This DAC-derived learning dynamics have been grounded in aspects of the physiology of the hippocampus [40] and pre-frontal cortex [32], and they reflect the core notions of Piaget's theory of cognitive development assimilation and accommodation through a process of equilibration [37, 56].

Human learners show a large variability in their performance and aptitude [16] requiring learning technologies to adjust to the skills and the progress of every individual. For learning to be efficient and applicable for as broad a range of students as possible, individual differences need to be taken into account. The critical condition that has to be satisfied, however, is that the *confusion* needs to be controllable so that it adjusts to the skills and the progress of individual students. This is consistent with the classical notion of Vygotsky's Zone of Proximal Development which is the level of knowledge that the learner can acquire with external assistance of a teacher or a peer [54]. Individualization thus serves the identification of this epistemic and motivational level.

Monitoring, controlling and adjusting the phase of confusion is what we call *shaping the landscape of success*. This approach is consistent to the notion of scaffolding, a technique based on helping the student to cross Vygotsky's Zone of Proximal Development. The concept of controlled confusion, as well as of individualized training, has already been tested in the context of neurorehabilitation using DAC based Rehabilitation Gaming System (RGS) which assists stroke patients in their functional recovery of motor deficits [10, 11]. RGS indeed effectively adjusts to individual users in terms of difficulty, allowing for an unsupervised deployment of individualized rehabilitation protocols.

Within the DAC architecture, the processes of learning are not isolated within single layers but they result as the interplay among them and the external world [51]. Although both the processes of learning deployed in the current experiment (resistance, confusion, resolution) and the layers of the DAC architecture (Reactive, Adaptive, Contextual) constitute a specific order and initial dependencies, their relation is not fixed. Depending on the learning goal (learning a new concept, contextualizing new information within a broader scale, etc.) the tutoring may be focusing on one of the three layers. In order to systematically traverse the three phases of learning distinguished here, the user is guided through a goal-based learning.

By incorporating DAC within the educational framework, our aim is to be able to create the feeling of resistance and confusion to introduce new knowledge specific for every individual student. Adjusting to the skills and the progress of individual students may result in keeping the process of acquisition motivating; so it is essential that despite helping the student to overcome certain difficulties, the task remains challenging enough.

3 THE SYNTHETIC TUTOR ASSISTANT (STA)

The STA emerges as the interplay of the three layers of DAC architecture. It is the STA that provides individualized content, adapted to the needs and capabilities of each student. Here we layout the framework for the implementation of the STA within the DAC architecture. The Reactive Layer provides the basic interaction between the student, tutor and teaching material through a self-regulation system and an allostatic control mechanism. It encompasses the basic reaction mechanisms guiding the student through the learning material in a predefined reactive manner and is based on a self-regulation mechanism that contains predefined reflexes that support behavior. Such reflexes are triggered by stimuli that can be either internal (self) or external (environment) and are coupled to specific affective states of the agent.

The Adaptive Layer will adjust the learning scenario to the needs and capabilities of the student based on the user model that is online updated throughout the analysis of the interaction. To do so, the STA needs to assess the state of the student (cognitive, physical, emotional), learn from previous interactions and adapt to each student. This knowledge will support the rich and multimodal interactions based on a the DAC control architecture. Finally, the Contextual Layer will monitor and adjust the learning strategy over long periods of time and over all participating students through Bayesian memory and sequence optimization. In the pilot experiment reported here, we are assessing the properties of the Reactive Layer of the STA in an educational scenario.

3.1 Behavioral modulation

In case of the STA, the main purpose of the self-regulating mechanism of the Reactive Layer is to provide the tutor with an initial set of behaviors that will initiate and maintain the interaction between the STA and the student. Grounded in biology, where living organisms are endowed with internal drives that trigger, maintain and direct behavior [25, 38], we argue that agents that are endowed with a motivational system show greater adaptability compared to simple reactive ones [2]. Drives are part of a homeostatic mechanism that aims at maintaining stability [12, 44], and various autonomous systems have used self-regulation mechanisms based on homeostatic regimes [6, 3].

Inspired by Maslow's hierarchy of needs [33], Hull's drive reduction theory [25] and tested in the autonomous interactive space Ada [15], the robots behavior is affected by its internal drives (for example the need to socialize - establish and maintain interaction). Each drive is controlled by a homeostatic mechanism. This mechanism classifies the drive in three main categories: *under*, *over* and *within* homeostasis. The main goal of the STA is to maximize its effectivity (or "happiness") as a tutor assistant, by maintaining its drives within specific homeostatic levels. To do so, the STA will need to take the appropriate actions. These states are focusing on the level of interaction with the learner and its consistency. Coherence at the behavioral level is achieved through an extra layer of control that reduces drives through behavioral changes, namely the allostatic control. Allostasis aims at maintaining stability through change [34]. The main goal of allostasis is the regulation of fundamental needs to ensure survival by orchestrating multiple homeostatic processes that directly or indirectly help to maintain stability.

The allostatic controller adds a number of new properties of the STA-DAC architecture, ensuring the attainment of consistency and balance in the satisfaction of the agent's drives and foundations for utilitarian emotions that drive communicative cues [53]. This approach strongly contradicts the paradigm of state machines standardly employed in comparable approaches and, in general, within the robotics community. State machines provide a series of closed-loop behaviours where each state triggers another state in function of its outcome. Here, drives are not associated on a one-to-one basis with a specific behavior. Instead, each behavior is associated with an intrinsic effect on the drives and with the usage of the allostatic controller, drives, and therefore behavior, change as the environment changes. With such design, drives modulate the robot's behavior adaptively in the function of every learner and the learning environment in general. Although in our current implementation, the mappings are hard-coded as reflexes (Reactive Layer), according to the DAC architecture, the mappings should be learnt through experience to provide adaptation.

3.2 The setup (software and hardware)

The DAC architecture and framework proposed are mostly hardware independent, as it can be applied in various robotic implementations [19, 42, 53, 31]. Here, the implementation aims at controlling the behavior of the robot and it involves a large set of sensors and effectors, designed to study Human-Robot Interaction (HRI). The setup (see figure 2) consists of the humanoid robot iCub (represented by the STA), the Reactable [23, 27] and a Kinect. The Reactable is a tabletop tangible display that was originally used as a musical instrument. It has a translucent top where objects and fingertips (cursors) are placed to control melody parameters. In our scenario, the usage of the Reactabe allows us to construct interactive games tailored to our needs. It furthermore provides information about the location of a virtual and physical object placed on the table and allows a precision that can hardly be matched using a vision based approach. In our lab, we have employed the Reactable in various interaction scenarios using the MTCF framework [28], such as musical DJ (cooperative game where the robot produces music with humans), Pong (competitive 2D simulated table tennis game) and Tic Tac Toe. The use and control of all these components allows the development of various interactive scenarios including educational games investigated here and allow the human and the robot to both act in a shared physical space. An extensive description of the overall system architecture can be found in [31, 52, 53]. The setup was designed to run autonomously in each trial, being the allostatic control the main component for providing the guidance for the learner/player during the task.

Figure 2. Experimental setup of the robot interacting with a human using the Reactable for the educational game scenario. In the image you can see the participant holding an object used to select an item from the Reactable (round table with projected images of countries and capitals). The participant is facing the iCub. The projected items are mirrored, so each side has the same objects.

4 TOWARDS ROBOTIC TEACHERS

In order to test the implementation of the STA-DAC as well as to evaluate the effectiveness of our scenario depending on different social features of the robot, we conducted a pilot study where the robot had the role of a tutor-peer.

The aim of the experiment focused on testing the effect of social cues (in this case, facial expression and eye contact) in HRI during an educational game. The goal was to test whether the variation of these social cues could affect the knowledge retrieval, subjective experience, and the very behavior towards the other player.

4.1 The educational scenario

The first question raised during the development of the STA is whether it can be an effective peer for the learner, both in terms of the social interactions and the impact on learning. Hence, the focus of this experiment is to study whether the modulation of certain behavioral parameters (based on the DAC architecture and the proposed behavioral modulation system), such as the use of eye contact and facial expressions, can change the acquisition of knowledge of a specific topic and the subjective experience of the user. On the one hand, eye contact can strengthen the interaction between the learner and the STA, for gazing can affect the knowledge transfer and the learning rate [36]. On the other hand, facial expressions can be used as a reinforcement of the participant's actions (the robot displays a happy face when the participant's choice is correct and a sad face when the matching is wrong), and could be considered as a reward.

The game-like scenario which we deployed is exercising Gagne's five learning categories [22]: verbal information, intellectual skill, cognitive strategy, motor skill and attitude. The game is based in a physical task, so the participants have to use their motor skills and, in order to solve the task, they have to develop a cognitive strategy to control their internal thinking processes. We also implemented three components of intellectual skill: concept learning, that is, learning about a topic; rule learning, used to learn the rules of the game; and, problem solving processes to decide how to match the pieces.

The educational scenario is a pairing game, where participants need to pair objects appearing on the Reactable to their corresponding categories. The pairing game is grounded in the premises of constructivism, where two or more peers learn together. Here the robot behaves similarly to a constructivist tutor: instead of just giving the information directly, it helps the student to understand the goal of the game (and, for example, reminding the subject the correct ways of playing) and it provides feedback regarding his actions (the robot only tells the correct answer to the subject when he has chosen a wrong answer, not before). For example, if the human selects a wrong pair, the robot indicates why the selection is wrong; it also comments on the correct selections. The players also receive visual information regarding their selection from the Reactable: if the selection is correct, the selected pair blinks with a green color and the object (but not the category) dissapears whereas the pair blinks with a red color if the selection is incorrect. The game was tested with both children and adults and the contents were adapted according their estimated knowledge. Therefore, for the children the game's topic was recycling, where the task was to correctly match different types of waste to the corresponding recycling bin. For the adults the topic was geography, where the task was to correctly match a capital with the corresponding country.

The learning scenario requires turn-taking and comprises three levels of increased difficulty. Both the human and robot had the same objects mirrored in each side. At each level, they had to correctly match the four objects to their corresponding category to proceed to the next level. The gradual increase of the difficulty allows for the scaffolding of the task, and consequently for the improvment of the learning process [4]. As mentioned earlier, the game was realized using the Reactable; the virtual objects were projected on the Reactable and object selection was achieved either with the usage of an object or with a cursor (fingertip). At the beginning of the interaction, the robot verbally introduces the game and is the first who initiates the interaction and the game.

4.2 Methods

We hypothesized that the combination of eye-contact and facial expressions strengthens the feedback between the player, the participant and the participant's choice, and affects the participant's subjective experience. As a result, we expected that when exposed to both behavioral conditions the participants would have a higher both knowledge transfer and the subjective experience.

To test our hypothesis and assess our architecture, we devised five experimental conditions (see Table 1) where we varied the gaze behavior and facial expressions of the STA. The experimental conditions are: Not-oriented Robot (NoR) (fixed gaze at a point - this way we are ensured that no eye contact is achieved); Task oriented Robot (ToR) (gaze supports actions, without making eye contact or showing facial expressions); Task and Human oriented Robot (T&HoR) (gaze supports actions, eye contact and showing facial expressions); Table-Human Interaction (THI), where the participant plays alone with the Reactable, and the Human-Human Interaction (HHI), where the participant plays with another human. Apart from the HHI, the behavior of the STA in terms of game play, verbal interaction and reaction to the participant's actions remained the same. The aim of the THI condition is to show the importance of embodiment of the STA during the interaction; the HHI condition acted as both the control group and a way of achieving a baseline regarding the interaction. The children were tested in the NoR, T&HoR and HHI conditions whereas the adults in all conditions.

Data were collected within three systems: knowledge and subjective experience questionnaires, behavioral data and the logs from the system. Participants had to answer pre- and post- knowledge questionnaires related to the pairing game. For recycling, the questionnaires had a total of twelve multiple-choice questions, including the same wastes and containers that the participants had to classify during the game. The information for creating this questionnaire came from the website "Residu on vas" (www.residuonvas.cat), property of the Catalan Wastes Agency. For geography, the questionnaires had a total of 24 multiple-choice questions (half of them, about the countries and capitals and the other half, about countries and flags). These questionnaires were given to the participants before and after the game, in order to evaluate their previous knowledge about the topic and later compare the pre- and post- knowledge results. The subjective experience questionnaire aims at assessing the STA's social behavior. It consists of 32 questions based on: the Basic Empathy Scale [26], the Godspeed questionnaires [5] and the Tripod Survey [17]. In the case of adults, there were 74 participants (age M = 25.18, SD = 7.55; 50 male and 24 female) distributed among five different conditions (THI=13, NoR=15, ToR=15, T&HoR=16, HHI=15). In the case of children, we tested 34 subjects (age M = 9.81, SD = 1.23; 23 male and 11 female) who randomly underwent three different experimental conditions (NoR=12, T&HoR=14, HHI=8).

Table 1. Table of the five experimental conditions.

	Embodiment	Action supporting gaze	Eye contact	Facial Expression
THI	No	No	No	No
NoR	Yes	No	No	No
ToR	Yes	Yes	No	No
T&HoR	Yes	Yes	Yes	Yes
HHI	Yes	Yes	Yes	Yes

Various conditions of robot behavior based on the interaction scenario

4.3 Results

First, we report a significant knowledge improvement in adults for all the conditions: THI, t(13) = 7. 697, p <0.001; NoR, (t(14) = 2.170, p = 0.048; ToR, t(14) = 3.112, p = 0.008, T&HoR, t(16) = 3.174, p = 0.006 and HHI,t(13) = 3.454 p = 0.004. In constrast, in children, there was no significance between conditions, although our results suggest a trend in improvement. We expected a difference among conditions, as we hypothesised that in the T&HoR condition, the knowledge transfer would be greater than the rest of the conditions. However this does not occur in neither the adult nor the children scenarios. In the case of children, we hypothesized that the associations were too simple; in the case of the adults, it seems that the knowledge transfer was achieved irregardless of the condition, suggesting that possibly the feedback of the Reactable itself regarding each pairing (green for correct and red for incorrect) might have been sufficient for the knowledge to be transferred.

Regarding the subjective experience, there was no statistical difference in the questionnaires data from children. We suspect that such result might be affected by the fact that both the Empathy and Godspeed questionnaires are designed for adults, and not children. In adults, although there was no significant difference among conditions for the Empathy and Tripod parts, there was a statistically significant difference between groups for the Godspeed part, as determined by one-way ANOVA (F(4,35) = 4.981, p = 0.003). As expected, humans scored higher (HHI, .06 $\pm$ 0.87), than the robot in two conditions (NoR, 2.84 $\pm$ 0.72, p = 0.003; ToR, 3.19 $\pm$ 0.46, p = 0.044, but surprisingly not in the T&HoR) and the table (THI,3.02 $\pm$ 0.56, p = 0.031) (Bonferroni post-hoc test). We can therefore hypothesize that the STA significantly scores lower than a human in all conditions but the one where its behavior is as close as possible to that of a human: gaze that sustains action (look at where the agent is about to point) and is used for communication purposes (look at human when speaking) and facial expressions as a feedback to the humans actions.

Regarding the behavioral data, there was a statistically significant difference between conditions for the mean gaze duration in children (one-way ANOVA (F(2,26) = 8.287, p = .0021)). A Bonferroni post-hoc test revealed that the time spent looking at the other player (in seconds) was significantly lower in the NoR (14.70 $\pm$ 8.81", p = 0.012) and the HHI conditions (11.74 $\pm$ 8.02", p = 0.003) compared to the T&HoR condition (30.97 $\pm$ 15.16")(figure 3). Our expectation regarding the difference between the NoR and T&HoR conditions was correctly met: people looked more at the agent who looked back at them. However, we were not expecting a difference between T&HoR and HHI condition. We believe that the reason why the difference in mean gaze duration occurs is because humans remained focused on the game and were mainly looking at table instead of looking at the other player. Furthermore, there were much less spoken interactions between them. In contrast, in the rest of the scenarios, the STA would comment on the actions of the participant, attrackting attention in more salient way.

In adults, a Kruskal-Wallis test showed that there was a high statistically significant difference in the time spent looking at the other player between the different conditions, x2(4) = 15.911, p = 0.003. The results of the Mann-Whitney Test showed significant differences between the THI (2.72 $\pm$ 5.53) and the NoR (16.37 $\pm$ 21.17) conditions (p = 0.026); the THI (2.72 $\pm$ 5.53) and the ToR (7.80 $\pm$ 7.76) conditions (p = 0.029); the THI (2.72 $\pm$ 5.53) and the T&HoR (19.87 $\pm$ 12.01) conditions (p <0.001); the ToR (7.80 $\pm$ 7.76) and the T&HoR (19.87 $\pm$ 12.01) conditions (p = 0.028); and the T&HoR

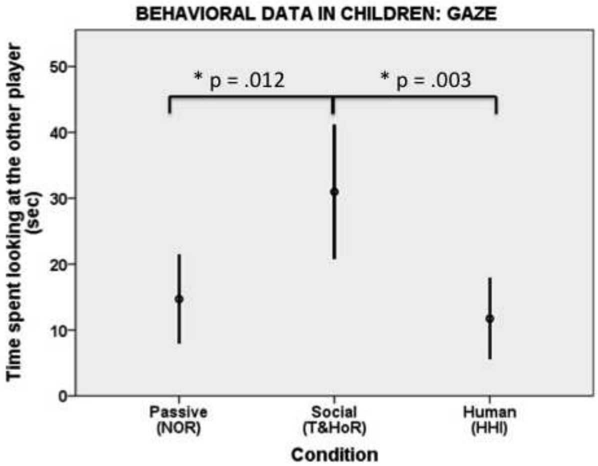


Figure 3. Time spent looking at the other player (in seconds) in children among conditions. Asterisks "*" depict significance.

(19.87 ± 12.01) and the HHI (3.66 ± 4.13) conditions (p = 0.002) (See figure 4). As expected, the more human-like the behavior of the STA, the more people would look at. The explanation regarding the difference between T&HoR and HHI in gaze duration is similar to that of children.

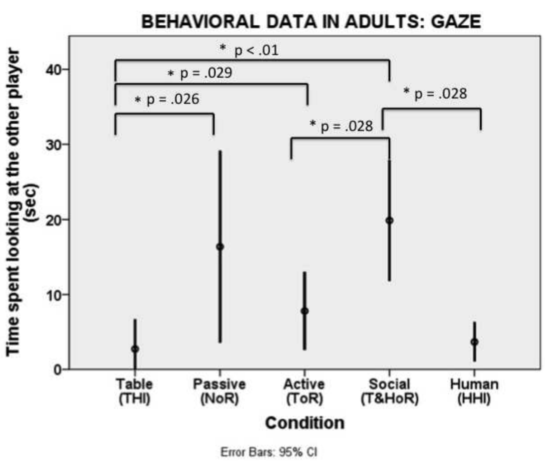


Figure 4. Time spent looking at the other player (in seconds) in adults among conditions. Asterisks "*" depict significance.

5 DISCUSSION AND CONCLUSIONS

The goal of the present study is to provide the key implementation features of the Synthetic Tutor Assistant (STA) based on the DAC architecture. Here, we propose the implementation of the STA within the DAC, a theory of the design principles which underlie perception, cognition and action. DAC is a layered architecture (Soma, Reactive, Adaptive and Contextual) intersected by three columns (world, self and actions), modeled to answer the H5W problem: Why, What, Where, When, Who and How. We explain the basic layers of DAC and focus on the Reactive Layer that constructs the basic reflexive behavioral system of the STA, as systhematically explained in section 3.1.

DAC predicts that learning is organized along a hierarchy of complexity and in order to acquire and consolidate new material the learner undergoes a sequence of learning phases: resistance, confusion and resolution. We argue that it is important to effectively adjust the difficulty of the learning scenario by manipulating the according parameters of the task (Adaptive Layer). This function will allow us for controlled manipulation of confusion, tailored to the needs of each student. Though it is not in the scope of the present study, in the future we plan to adjust the parameters of the learning scenario studied here on the basis of an online analysis of the learners' performance, interpreted both in terms of traditional pedagogical scales and the DAC architecture (Adaptive Layer). The learner's errors and achievements will be distinguished in terms of specific hierarchical organization and dynamics. Finally, the Contextual Layer will monitor and adjust the difficulty parameters for both individual students and bigger groups on a longer time scales. The motivational system presented is mainly focused on the Reactive Layer of the architecture, but our aim is to primarily adapt the Reactive Layer to the needs of STA and teaching scenarios and then extend the STA to include the Adaptive and Contextual Layers.

We devised an educational scenario to test the implementation of the STA-DAC as well as to evaluate the effectiveness of different social features of the robot (social cues such as eye contact and facial expressions). The task devised was a pairing game using the Reactable as an interface, where the robot acts as a constructivist tutor. The pairing consisted of matching different types of waste to the corresponding recycling bin (recycle game) for the children and matching the corresponding capital to a country (geography game) for the adults. The learning scenario was turn-taking with three levels of increased difficulty. The experiment consists of five different conditions, described in section 4.2: THI, NoR, ToR, T&HoR and HHI. Adults were tested in all conditions whereas children in NoR, T&HoR and HHI. To assess the interaction, the implementation as well as the effectiveness of the robot's social cues, behavioral data, logged files and questionnaires were collected.

In the results, we see that in adults, there are significant differences in knowledge improvement among conditions. On the other hand, there is a trend in knowledge improvement in children, but it is not significant. The results are not sufficient to draw any concrete conclusions about knowledge retrieval. Nevertheless, we can see that people scored higher in the post-experiment questionnaire, on the other hand, results are not enough to identify exactly the reason. It is possible that the task, though the difficulty increased on each trial, would still remain relatively easy. That is why we aim at devising a related experiment where we would exploit the Adaptive Layer that adapts the difficulty to each individual player.

Our results show that children looked more at the T&HoR robot than then ToR or HHI. Based on these results, we can conclude that the behavior of the Task and Human oriented Robot drew more the attention of the participant than the other human or the solely Task oriented Robot. The robot was looking at the participant when it was addressing him; its gaze followed both the player's and its own actions, meaning that it would look at the object that the participant had chosen or the object that it chose. Finally, it would show facial expressions according to each event: happy for the correct pair or sad for the incorrect one. Such cues may indeed be more salient and draw the attention of the player. In all conditions, the robot was speaking, so it seems that it was the implicit non-verbal communicative signals of the robot that drew the attention of the participant. In the case of the adults, the results are also similar. Such behavior is important in the development of not only social but also educational robots, as gaze following directs attention to areas of high information value and accelerates social, causal, and cultural learning [1]. Indeed, such cues positively impact human-robot task performance with respect to understandability [7]. This is supported by results like the ones of [24], where the addition of gestures led to a higher effect on the participant only when the robot was also performing eye contact.

Finally, the results from the Godspeed questionnaire in adults show a significant difference in the overall score between HHI and THI, NoR, ToR but not the T&HoR. Such results were generally expected, as a human would score higher than the machine. In children, there was no significance in any of the conditions, however, it may be the case that the Godspeed questionnaire is not the optimal measurement for subjective experience, at it may contain concepts that are not yet fully understood by such a young age. Perhaps simpler or even more visual (with drawings that represent the extremes of a category) questionnaires would be more appropriate.

Though the knowledge transfer results are not sufficient to draw any concrete conclusions (as the knowledge transfer is not significantly different among conditions), the complex social behavior of the robot indeed attracts attention of the participant. As for the pilot study, the authors need to focus more on the evaluation of the system, and need to introduce a strong experimental design to derive more specific conclusions. Further analysis of the behavioral data can provide insight regarding eye contact in terms of error trials, decision time and task difficulty. In the upcoming experiments we will provide a better control in the HHI condition. A possible strategy is to deploy a specific person (an actor) as the other player, to normalize the characteristics of the scenario between all the subjects.

ACKNOWLEDGEMENTS

This work is supported by the EU FP7 project WYSIWYD (FP7-ICT-612139) and EASEL (FP7-ICT- 611971).

REFERENCES

[1] '"Social" robots are psychological agents for infants: a test of gaze following.', *Neural networks : the official journal of the International Neural Network Society*, **23**(8-9), 966–72, (2010).
[2] Michael A Arbib and Jean-Marc Fellous, 'Emotions: from brain to robot', *Trends in cognitive sciences*, **8**(12), 554–561, (2004).
[3] Ronald C Arkin, Masahiro Fujita, Tsuyoshi Takagi, and Rika Hasegawa, 'An ethological and emotional basis for human–robot interaction', *Robotics and Autonomous Systems*, **42**(3), 191–201, (2003).
[4] Roger Azevedo and Allyson F Hadwin, 'Scaffolding self-regulated learning and metacognition–implications for the design of computer-based scaffolds', *Instructional Science*, **33**(5), 367–379, (2005).
[5] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi, 'Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots', *International journal of social robotics*, **1**(1), 71–81, (2009).
[6] Cynthia Breazeal, 'Emotion and sociable humanoid robots', *International Journal of Human-Computer Studies*, **59**(1), 119–155, (2003).
[7] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin, 'Effects of nonverbal communication on efficiency and robustness in human-robot teamwork', in *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pp. 708–713. IEEE, (2005).
[8] Cynthia L Breazeal, *Designing sociable robots*, MIT press, 2004.
[9] Sylvain Calinon, Danilo Bruno, Milad S Malekzadeh, Thrishantha Nanayakkara, and Darwin G Caldwell, 'Human–robot skills transfer interfaces for a flexible surgical robot', *Computer methods and programs in biomedicine*, (2014).
[10] Mónica S Cameirão, Sergi B Badia, Esther Duarte Oller, PFMJ Verschure, et al., 'Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation', *Journal of neuroengineering and rehabilitation*, **7**(1), 48, (2010).
[11] Mónica S Cameirão, Sergi Bermúdez i Badia, Esther Duarte, Antonio Frisoli, and Paul FMJ Verschure, 'The combined impact of virtual reality neurorehabilitation and its interfaces on upper extremity functional recovery in patients with chronic stroke', *Stroke*, **43**(10), 2720–2728, (2012).
[12] Walter Bradford Cannon, 'The wisdom of the body.', (1932).
[13] Paolo Dario, Paul FMJ Verschure, Tony Prescott, Gordon Cheng, Giulio Sandini, Roberto Cingolani, Rüdiger Dillmann, Dario Floreano, Christophe Leroux, Sheila MacNeil, et al., 'Robot companions for citizens', *Procedia Computer Science*, **7**, 47–51, (2011).
[14] Armin Duff and Paul FMJ Verschure, 'Unifying perceptual and behavioral learning with a correlative subspace learning rule', *Neurocomputing*, **73**(10), 1818–1830, (2010).
[15] Kynan Eng, David Klein, A Babler, Ulysses Bernardet, Mark Blanchard, Marcio Costa, Tobi Delbrück, Rodney J Douglas, Klaus Hepp, Jonatas Manzolli, et al., 'Design for a brain revisited: the neuromorphic design and functionality of the interactive space 'Ada'', *Reviews in the Neurosciences*, **14**(1-2), 145–180, (2003).
[16] RM Felder and R Brent, 'Understanding student differences', *Journal of engineering education*, **94**(1), 57–72, (2005).
[17] R Ferguson, 'The tripod project framework', *The Tripod Project*, (2008).
[18] Ylva Fernaeus, Maria Håkansson, Mattias Jacobsson, and Sara Ljungblad, 'How do you play with a robotic toy animal?: a long-term study of pleo', in *Proceedings of the 9th international Conference on interaction Design and Children*, pp. 39–48. ACM, (2010).
[19] Marti Sanchez Fibla, Ulysses Bernardet, and Paul FMJ Verschure, 'Allostatic control for robot behaviour regulation: An extension to path planning', in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, pp. 1935–1942. IEEE, (2010).
[20] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn, 'A survey of socially interactive robots', *Robotics and autonomous systems*, **42**(3), 143–166, (2003).
[21] Marina Fridin, 'Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education', *Computers & Education*, **70**, 53–64, (2014).
[22] Robert Mills Gagné, *The conditions of learning and theory of instruction*, Holt, Rinehart and Winston New York, 1985.
[23] Günter Geiger, Nadine Alber, Sergi Jordà, and Marcos Alonso, 'The reactable: A collaborative musical instrument for playing and understanding music', *Her&Mus. Heritage & Museography*, (4), 36–43, (2010).
[24] J. Ham, B. René, R. Cuijpers, D. van der Pol, and J. J. Cabibihan, 'Making Robots Persuasive: The Influence of Combining Persuasive Strategies (Gazing and Gestures) by a Storytelling Robot on Its Persuasive Power', in *Third International Conference, ICSR 2011 Amsterdam, The Netherlands, November 24-25, 2011 Proceedings*, pp. 71–83, (2011).
[25] Clark Hull, 'Principles of behavior', (1943).
[26] Darrick Jolliffe and David P Farrington, 'Development and validation of the basic empathy scale', *Journal of adolescence*, **29**(4), 589–611, (2006).
[27] Sergi Jordà, 'On stage: the reactable and other musical tangibles go real', *International Journal of Arts and Technology*, **1**(3), 268–287, (2008).
[28] Carles F Julià, Daniel Gallardo, and Sergi Jordà, 'Mtcf: A framework for designing and coding musical tabletop applications directly in pure data', in *Proceedings of the International Conference on New Interfaces for Musical Expression*, volume 20011, (2011).
[29] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro, 'Interactive robots as social partners and peer tutors for children: A field trial', *Human-computer interaction*, **19**(1), 61–84, (2004).
[30] Cory D Kidd and Cynthia Breazeal, 'Robots at home: Understanding long-term human-robot interaction', in *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, pp. 3230–3235. IEEE, (2008).
[31] Stéphane Lallée, Vasiliki Vouloutsi, Maria Blancas Munoz, Klaudia Grechuta, Jordi-Ysard Puigbo Llobet, Marina Sarda, and Paul FMJ Verschure, 'Towards the synthetic self: making others perceive me as an other'.
[32] Encarni Marcos, Pierpaolo Pani, Emiliano Brunamonti, Gustavo Deco,

Stefano Ferraina, and Paul Verschure, 'Neural variability in premotor cortex is modulated by trial history and predicts behavioral performance', *Neuron*, **78**(2), 249–255, (2013).
[33] Abraham Harold Maslow, 'A theory of human motivation.', *Psychological review*, **50**(4), 370, (1943).
[34] Bruce S McEwen, 'Allostasis and allostatic load: implications for neuropsychopharmacology', *Neuropsychopharmacology*, **22**(2), 108–124, (2000).
[35] Seymour Papert, *Mindstorms: Children, Computers, and Powerful Ideas*, Basic Books, Inc., New York, NY, USA, 1980.
[36] Fiona G Phelps, Gwyneth Doherty-Sneddon, and Hannah Warnock, 'Helping children think: Gaze aversion and teaching', *British Journal of Developmental Psychology*, **24**(3), 577–588, (2006).
[37] Jean Piaget and Margaret Trans Cook, 'The origins of intelligence in children.', (1952).
[38] Robert Plutchik, *The emotions*, University Press of America, 1991.
[39] Tony J Prescott, Nathan Lepora, and Paul FMJ Vershure, 'A future of living machines?: International trends and prospects in biomimetic and biohybrid systems', in *SPIE Smart Structures and Materials+ Nondestructive Evaluation and Health Monitoring*, pp. 905502–905502. International Society for Optics and Photonics, (2014).
[40] César Rennó-Costa, John E Lisman, and Paul FMJ Verschure, 'A signature of attractor dynamics in the ca3 region of the hippocampus', *PLoS computational biology*, **10**(5), e1003641, (2014).
[41] Martin Saerbeck, Tom Schut, Christoph Bartneck, and Maddy D Janse, 'Expressive robots in education: varying the degree of social supportive behavior of a robotic tutor', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1613–1622. ACM, (2010).
[42] Marti Sanchez-Fibla, Ulysses Bernardet, Erez Wasserman, Tatiana Pelc, Matti Mintz, Jadin C Jackson, Carien Lansink, Cyriel Pennartz, and Paul FMJ Verschure, 'Allostatic control for robot behavior regulation: a comparative rodent-robot study', *Advances in Complex Systems*, **13**(03), 377–403, (2010).
[43] Brian Scassellati, 'Using social robots to study abnormal social development', (2005).
[44] John P Seward, 'Drive, incentive, and reinforcement.', *Psychological review*, **63**(3), 195, (1956).
[45] Gerry Stahl, 'What we know about cscl and implementing it in higher education', chapter Building Collaborative Knowing: Elements of a Social Theory of CSCL, 53–85, Kluwer Academic Publishers, Norwell, MA, USA, (2004).
[46] PAUL VERSCHURE, 'Formal minds and biological brains ii: from the mirage of intelligence to a science and engineering of consciousness', (2013).
[47] Paul FMJ Verschure, 'Distributed adaptive control: A theory of the mind, brain, body nexus', *Biologically Inspired Cognitive Architectures*, (2012).
[48] Paul FMJ Verschure and Philipp Althaus, 'A real-world rational agent: Unifying old and new ai', *Cognitive science*, **27**(4), 561–590, (2003).
[49] Paul FMJ Verschure, Cyriel MA Pennartz, and Giovanni Pezzulo, 'The why, what, where, when and how of goal-directed choice: neuronal and computational principles', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1655), 20130483, (2014).
[50] Paul FMJ Verschure, Cyriel MA Pennartz, and Giovanni Pezzulo, 'The why, what, where, when and how of goal-directed choice: neuronal and computational principles', *Philosophical Transactions of the Royal Society B: Biological Sciences*, **369**(1655), 20130483, (2014).
[51] Paul FMJ Verschure, Thomas Voegtlin, and Rodney J Douglas, 'Environmentally mediated synergy between perception and behaviour in mobile robots', *Nature*, **425**(6958), 620–624, (2003).
[52] Vasiliki Vouloutsi, Klaudia Grechuta, Stéphane Lallée, and Paul FMJ Verschure, 'The influence of behavioral complexity on robot perception', in *Biomimetic and Biohybrid Systems*, 332–343, Springer, (2014).
[53] Vasiliki Vouloutsi, Stéphane Lallée, and Paul FMJ Verschure, 'Modulating behaviors using allostatic control', in *Biomimetic and Biohybrid Systems*, 287–298, Springer, (2013).
[54] Lev S Vygotsky, *Mind in society: The development of higher psychological processes*, Harvard university press, 1980.
[55] Kazuyoshi Wada and Takanori Shibata, 'Living with seal robots in a care house-evaluations of social and physiological influences', in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 4940–4945. IEEE, (2006).
[56] Barry J Wadsworth, *Piaget's theory of cognitive and affective development: Foundations of constructivism* ., Longman Publishing, 1996.