Useful and Motivating Robots – The Influence of Task Structure on Human-Robot Teamwork

Manja Lohse
Human Media Interaction
University of Twente
Enschede, the Netherlands
m.lohse@utwente.nl

Vanessa Evers
Human Media Interaction
University of Twente
Enschede, the Netherlands
v.evers@utwente.nl

ABSTRACT

Robots have recently started to leave their safety cages to be used in close vicinity to humans. This also causes changes in the nature of the tasks that robots and humans solve together, i.e., in the degree of structure of the tasks. While traditional, industrial tasks were highly structured, the new tasks often have a low level of structure. We present a user study that compares a highly and a little structured task in a text-based computer game played by human-robot teams. The results suggest that users do not only find robots useful and motivating in highly structured tasks where they depend on their help, but also in little structured tasks that they could solve on their own.

Categories and Subject Descriptors

H.m [**Information Systems**]: Miscellaneous

General Terms

Experimentation

Keywords

human-robot teamwork, task structure

1. INTRODUCTION

Within the last decades, teamwork has been recognized as an effective means of conducting business. Teamwork has also advanced collaborative work environments where humans and robots work together on shared tasks, e.g., in surgery or in industry where humans and robots work jointly on assembling complex products such as cars. Thus, robots leave their safety cages to interact in human-robot teams [2, 7]. The nature of teamwork is changing through the introduction of such autonomous robots and the topic is starting to receive attention in HRI [2, 3, 7]. One aspect of teamwork that has not been explored to date is the degree of task structure. By task structure we refer to characteristics of tasks such as how well defined the goals of the tasks are (goal clarity), how many ways there are to achieve the goals (goal path multiplicity), how well the actors know whether their steps to achieve the goals were correct (decision verifiability), and how high the level of interdependence between the actors involved in the task is (task interdependence) [4]. Task structure is highly relevant because the new robotic applications mentioned above require largely unstructured, non-routine tasks compared to the highly structured, routine tasks in factory settings. Unstructured tasks are characterized by little defined goals, many ways to achieve them, little possibilities to check whether decisions were optimal, and little interdependence between the actors. To date there is no research on the question whether users accept robots to assist also in such tasks or only in highly structured ones. To fill this gap, we investigated the influence of the task structure on human-robot teamwork in a text-based computer game. We compared a highly and a little structured task in order to determine how task structure influences the perception of a robot's usefulness, ability to motivate the users, and further criteria of user experience. Furthermore, we analyzed whether the users' behavior objectively differed between the conditions with respect to how much effort they put into playing the game and what commands they used.

HRI'14, March 3–6, 2014, Bielefeld, Germany.
ACM 978-1-4503-2658-2/14/03.
http://http://dx.doi.org/10.1145/2559636.2559803.

2. SYSTEM DESIGN

Our goal was to implement a highly and a little structured task in a way that allows to compare them without having to consider many differences that are due to the task design instead of the degree of structure of the task. Therefore, we decided to implement a text-based computer game. The goal of the game was to find the statue of a golden goose and to become a rich explorer. The players could move around between screens, pick up objects, use them to access new screens, and solve riddles. The game was designed in a way that offered several opportunities for a robot to engage the player in teamwork such as providing the player with information about possible destinations, objects, and riddles. We implemented these behaviors in a Nao robot using the Choregraphe software. Nao behaviors were executed either when the user asked the robot for help (verbal triggers), when the game entered a certain state, e.g. a new screen, (game triggers), or when the researcher had to intervene (manual command triggers).

The task structure of the game was manipulated based on the factors introduced above. *Goal clarity* was manipulated in the instructions that the participants received before the interaction with the robot (goal clearly stated vs. rather

vague goal statement). Different degrees of *goal path multiplicity* were achieved by implementing the highly structured version of the game in a way that only one path led to the goal, and multiple paths led to the goal in the little structured version. These differences also affected the *solution specificity*. While there was one clear solution in the highly structured version (reach the golden goose), multiple solutions to collect points were available in the little structured version of the game. A high degree of *decision verifiability* was realized in the game by providing specific feedback to the consequences of actions in the highly structured version and only ambiguous feedback in the little structured version. All these criteria also affected the *interdependence* between the actors. Asking the Nao for help to complete the game was required only in the highly structured version, thus, increasing the interdependence between human and robot. While the game differed between the conditions, the Nao behaviors largely stayed the same.

3. SAMPLE

23 participants took part in our user study (6 female, 17 male; aged 19 to 31, with a mean age of 23.3 years, sd = 2.50). All participants were students at the University of Twente. They were divided in two groups (one per condition). Both groups included the same number of female participants and had roughly the same age distribution. The participants had little experience with text based games (m = 2.45; sd = 1.30)(all items on a scale of 1: not at all - 5: a lot) and even less interacting with the Nao robot (m = 1.24, sd = 0.77).

4. RESULTS

We measured *user experience* with the USE scales by Lund (cited from [5]). All items were rated on a 5-point Likert scale (1: strongly disagree - 5: strongly agree). They were presented in random order. The scale did not reveal any significant differences regarding the robot between the two conditions. The *usefulness* (m = 3.85; sd = 0.59); (Cronbach's α .820) and the *satisfaction* (m = 3.80; sd = 0.50); (Cronbach's α .729) were experienced very positively in both conditions and also the *ease of use* (m = 3.56; sd = 0.55); (Cronbach's α .716) had a tendency toward the positive side of the scale. We employed the two one-sided test (TOST) to determine whether the user experience of the robot was actually equivalent in both conditions [6]. The equivalence margin (δ) was set to the standard deviations of the individual subscales. Following the TOST procedure, the two conditions could be declared equivalent for all subscales; for each of the dependent variables the lower and upper bound were within the mean difference $\pm\delta$ margin.

In contrast to the user experience of the robot, we found a non-significant tendency that the game was easier to use in the highly structured condition (m = 3.52; sd = 0.53) than in the little structured condition (m = 3.11; sd = 0.45); (T(19)=1.882; p=.075; two-tailed); (Cronbach's α .726). However, the users' *satisfaction* with the game was significantly higher in the highly structured condition (m = 3.92; sd = 0.64) than in the little structured version (m = 3.20; sd = 0.58); (T(19)=2.633, p=.016; two-tailed); (Cronbach's α .813). Overall, these results point to a more positive user experience of the highly structured game.

Additionally, we measured the *robot's capability to motivate users* with four items (Cronbach's α = .696) inspired by the work of Fasola and Matarić [1]. These ratings did not reveal any differences between the conditions. Overall the Nao worked well as a motivator (m=3.63; sd = 0.70). Again, we employed the TOST procedure as described above and the two conditions could be declared equivalent.

We analyzed the logfiles to determine whether the *actions that people took* in the game differed between the conditions. Overall, we found that the time to play the game was highly similar in both conditions (highly structured: m = 22.50 minutes, sd = 5.23; little structured: m = 23.89 minutes, sd = 1.71). Also the mean number of actions was similar (highly structured: m= 220.00, sd = 68.66; little structured: m = 250.11, sd = 65.00). The TOST procedure showed that the two conditions could be declared equivalent. The analysis of the logfiles also revealed that participants used most commands equally often in both conditions.

5. CONCLUSION

Our results show that the robot was perceived as a useful and motivating team member in both a highly and a little structured task. While we expected that the robot would be perceived as even more useful in the highly structured task which involved higher interdependence between human and robot, this was not the case. Also the satisfaction with the robot and the ease of use were rated similarly and the robot was perceived as being equally motivating in both conditions. This finding is very encouraging for our research, particularly given the fact that the game was perceived as being easier to use and more satisfying in the highly structured version. Thus, the robot succeeded as a member of the team in the little structured task despite the fact that it involved less interdependence and a more negative evaluation of the game. We conclude, that users indeed find robots that support little structured tasks useful and motivating.

6. REFERENCES

[1] J. Fasola and M. J. Matarić. Using socially assistive hri to motivate physical exercise for older adults. *Proc. of the IEEE*, 100(8):2512 – 2526, 2012.

[2] G. Hoffman and C. Breazeal. Effects of anticipatory perceptual simulation on practiced human-robot tasks. *Autonomous Robots*, 28(4):403–423, 2010.

[3] J. Shah and C. Breazeal. An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors: The Journal of the Human*, 10(2):234–245, 2010.

[4] M. Shaw. Scaling group tasks: A method for dimensional analysis. Technical report, University of Florida, 1963.

[5] T. Tullis and W. Albert. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2008.

[6] E. Walker and A. S. Nowacki. Understanding equivalence and noninferiority testing. 26(2):192–196, 2011.

[7] R. Wilcox, S. Nikolaidis, and J. Shah. Optimization of temporal dynamics for adaptive human-robot interaction in assembly manufacturing. *Proceedings of Robotics: Science and Systems*, 2012.