# Groups in Conflict at the Airport: How People Think a Robot Should Act

**Michiel Joosse**
Human Media Interaction group
University of Twente
Enschede, the Netherlands
m.p.joosse@utwente.nl

**Sanne van Waveren**
Human Media Interaction group
University of Twente
Enschede, the Netherlands
s.vanwaveren@student.utwente.nl

**Cristina Zaga**
**Vanessa Evers**
Human Media Interaction group
University of Twente
Enschede, the Netherlands
c.zaga@utwente.nl
v.evers@utwente.nl


*CSCW'17 Workshop on Robots in Groups and Teams*, 26 February 2017, Portland, Oregon.

## Abstract
As robots provide services to groups of people conflicts can arise forcing the robots to decide which action to take. In this extended abstract we present a study in which we collected qualitative data on peoples' ideas of appropriate robot responses when encountering conflict situations while guiding passengers at an airport. We collected, coded and analyzed free text data from 118 participants, and will report on the dataset and general results. In general participants' believed the needs of the many outweigh the needs of the few: the robot should move on.

## Author Keywords
Guide robots; behavior; cultural differences; crowdsourcing

## ACM Classification Keywords
H.1.2 [Information Systems]: Human factors

## Introduction
Social robots are envisioned to provide services in a variety of environments; ranging from robots in the classroom to robots providing autonomous tours in zoos [3] and other public spaces. In order to deploy robots in public space there are a number of challenges; both technical and social in order to create robots which are safe and at the same time behave in a socially acceptable way.

**Instruction:**
*"Please imagine yourself being at an unknown airport where you transfer from one flight to another. The robot in the picture guides you and 2 other passengers, whom you haven't met before, to the same gate where the connecting flight will depart. You have more than enough time to walk to your connecting flight. Please imagine the following scenario:*

*The minority of other group members falls behind and can't keep up with the speed of the robot guide and the rest of the group."*

In Human-Robot Interaction (HRI), user studies investigating appropriate robot behaviors cover a variety of topics, such as socially normative navigation [7], approaching individual people [10] and groups [9], and cultural differences in human-robot approach distances [2]. The current work concerns the SPENCER project, an international project in which we develop and deploy a robot to guide small groups of transfer passengers at an airport from an arrival gate to their departure gate [8]. These groups are envisioned to be composed of one or more subgroups who, besides their arrival- and departure flight, have nothing in common. From literature in organizational management, we know that subgroups are more likely to get into conflict when there are one or more (demographic or organizational) faultlines present [6]. We feel that it is relevant to investigate which actions a robot should take when conflict arises within groups. As such, this contribution is mostly related to the question of how a robot shapes the dynamics of groups and teams in existing settings.

In this paper, we report on exploratory study in which we investigated what people think a robot should do when conflict arises; for example in cases where a part of the group urgently needs to use the bathroom. We collected and coded qualitative data free text) from 118 participants who described how a robot should act in a variety of situations. We will contribute to the design of robot behaviors in order to participate as group members. We will report on the methodology used to collect the data, our data analysis process and initial results. We will conclude with a discussion of the results.

## Data collection

We designed a survey containing descriptions of situations which provide a conflict for a robot guide, and which are likely occur at an airport. This survey was distributed to people living in the United States and China as previous works found cultural differences between these two (national) cultures in their behavior towards robots [11], attitude towards robots [1] and preferences for behaviors [5]. Based upon a contextual analysis of an international airport [5] and common knowledge of transferring we defined 10 conflicts which could occur when transferring:

1. Group members cannot keep up with the robot.
2. People forgot their passport (or other important documents), and thus have to return to the collect them.
3. A portion of the group enters a store to buy a present.
4. Group members urgently have to go to the bathroom.
5. An elderly person, not part of the group, crosses the path of the robot, requiring the robot to slow down.
6. A member of the group stops to tie his/her shoelace.
7. A member of the group is lost, and cannot find the robot.
8. One of the group members stand in the way of the robot to take a selfie.
9. One of the group members wants to have a smoke.
10. One of the group members meets someone familiar, stops and starts a conversation.

*Manipulations*

Within the survey we manipulated four variables, the first two being *time-pressure* and the *size of the group* being guided. We manipulated time pressure as we found in our previous work that people behave differently in an experiment when they are primed with time pressure using simple verbal instructions. We manipulated group size as we believe people who are part of a smaller group are more likely to have the robot wait in certain situations. The third and fourth variables we manipulated concerned the applicability of the social situation. The situation concerned either

| | Code | % |
|---|---|---|
| 1 | Continue | 25.9% |
| 2 | Slow down | 6.0 % |
| 3 | Wait | 31.0% |
| 4 | Assist person | 3.2% |
| 5 | Attract attention | 1.9% |
| 6 | Go with group | 3.5% |
| 7 | Provide information | 8.0% |
| 8 | Involve other | 3.5% |
| 9 | (other answer) | 8.4% |
| 10 | (answer unclear) | 8.6% |

**Table 1:** Data was categorized into 1 of 10 categories. Two coders coded 30% of the data ($\kappa$=0.705). Most data was coded as "Wait"

the *minority or majority* of the group, and the participant filling out the questionnaire was also *involved or not involved* in the problem (e.g. part of the group who has to go to the bathroom). We therefore had (2x2x2=)8 conditions, for each of the two nationalities (China and the United States). As stated above, participant involvement in the cause of situation was manipulated within-subjects, with the exception of situation 5, therefore all participants were provided with (20-1=)19 situations. Because this introduces a violation of assumptions for the chi-square tests used in the analysis, we did not use all data collected in the analysis; rather only the situations in which the participant was not involved in the cause of the conflict.

*Materials*
We collected data through a web-based survey. The survey started with an overview of the survey, and technical data of the SPENCER robot, including the height and weight, capabilities and interaction modalities. Following this information, participants were provided with the situation at hand. An example is provided in the sidebar on the previous page. The survey also contained questions about participants' experience with robots and demographic questions. This resulted in a survey containing a total of 66 questions. The survey itself was hosted on SurveyMonkey, and participant recruitment and payment was conducted through the CrowdFlower platform.

*Participants*
The survey was distributed to participants in China and the United States. A total of 225 participants completed the survey. We excluded 107 participants (47%) who satisfied one or more of the following three exclusion criteria:

1. Participants whose answers did not consist of text (words, sentences), but rather random numbers or meaningless strings of characters.

| | Time pressure | Group size | Applicability | US | China | Total |
|---|---|---|---|---|---|---|
| A | high | small | minority | 9 | 10 | 19 |
| B | high | small | majority | 11 | 9 | 20 |
| C | high | larger | minority | 8 | 5 | 13 |
| D | high | larger | majority | 8 | 5 | 13 |
| E | low | small | minority | 6 | 5 | 11 |
| F | low | small | majority | 11 | 7 | 18 |
| G | low | larger | minority | 7 | 5 | 12 |
| H | low | larger | majority | 6 | 6 | 12 |
| | | | **Total** | 66 | 52 | 118 |

**Table 2:** Participant distribution over conditions

2. Participants who did not provide complete sentences, or the same answer for all 19 situations)
3. Participants who failed to answer the manipulation check question.

Data of 118 participants was included in the data analysis. Table 2 provides information on the distribution of participants in each condition. A total of 72 males and 45 females participated in the study, aged between 17 and 66 (*M*=32.33, *SD*=11.99). The sample had a diverse background and consisted of students (16.9%), unemployed people (11.0%), and people with diverse occupational backgrounds such as education (11.0%), construction (5.1%), health care (5.9%) and finance (5.1%).

## Data analysis & dataset descriptives

The dataset contained a total of 1180 answers[1], ranging from short answers of 4 characters to stories of 255 char-

[1] 118 participants each completed 10 questions

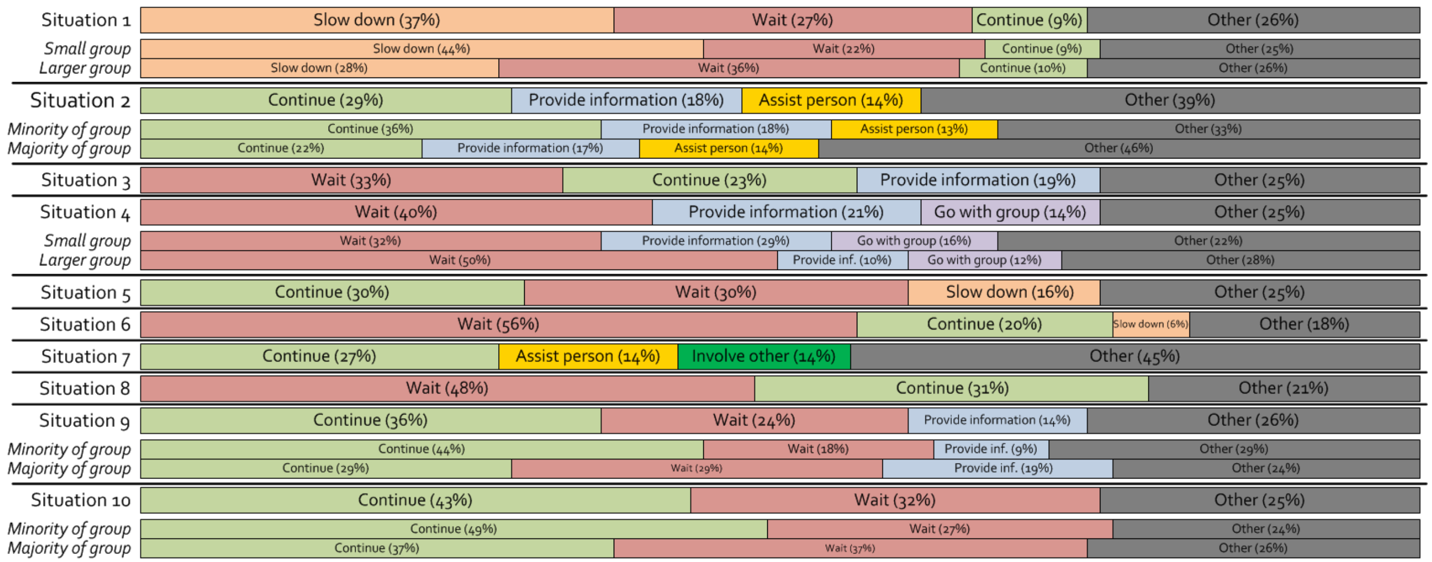


**Figure 1:** Top-3 of coded answers, per question. Where relevant divisions have been made on e.g. group size or situation applicability.

**Examples of answers provided by participants**

“Making some music to reminder the group member to be able to catch up with team."
*(Situation 10)*

“Robot should stop and wait for group. It Should also signal to the destination gate that the group is en route. Perhaps signal store to issue a hurry-up request to group."
*(Situation 3)*

acters. The answers were coded by two coders (31% of the data was coded by both coders), and each answer was initially assigned one or multiple codes out of a potential 27. As it proved impossible to provide a meaningful analysis of these data, it was eventually decided to narrow the analysis, and group a number of codes into 1 of 10 categories. We then assessed intercoder reliability based upon that category, which was moderately-high ($\kappa$=0.705).

Before presenting the results, we will first present some characteristics of the coded dataset. Most of the data was coded in either the “wait" (31.0%) or “continue" (25.9%) categories (Table 1). Roughly 16% of the data could not be coded into one of eight categories describing actions for the robot. Data from female participants was significantly less coded as “other" ($z$=-2.0) and “unclear" ($z$=-2.4); thus being easier to interpret.

We ran a series of chi-square tests on the categorical data, and used the standardized residuals to test if the data differed significantly from the expected value [4, p. 699]. For example: if participants in (f.e.) the small group were significantly more likely to have the robot wait, as compared to participants in the larger group. None of the standardized residuals yielded significant results. We believe the descriptive data provides directions and input for future research, which we will present in the results section and reflect upon in the discussion section.

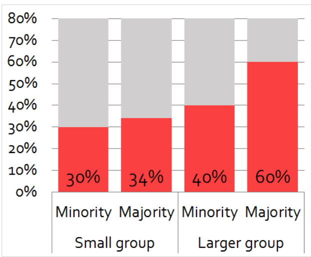


**Figure 2:** Especially when the majority of a large group has to go to the bathroom, participants believe a guide robot should wait.

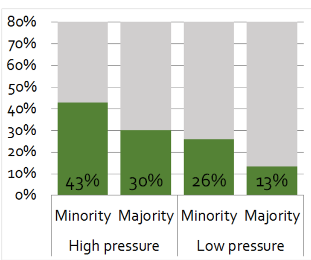


**Figure 3:** As time pressure increases, and the situations applies to less people participants are more likely to have the robot continue

## Results

For the purposes of this workshop we took a look at the descriptive data and we visualized the most relevant results in Figure 1. The way the data is reported is as follows: we report the percentages of the top-3 categories in which the most answers were coded, which are sorted from left-to-right following a majority voting principle. Where relevant, we have also included a division based upon one of the manipulations. This is not an exhaustive list of the various differences within the conditions, however, we feel these are the most interesting for the purposes of the workshop.

*Let's move along now*

In 6 out of 10 situations most of the participants indicated the robot should continue or perhaps slow down. Two of four situations where the robot is supposed to wait are actions which do not necessarily have to take up a lot of time; these are the situations where a group members ties his or her shoelace and when someone takes a selfie. Though we have not asked participants why they would consider an action appropriate, it could be that in the other two cases (buy a present and go to the bathroom) participants might have considered these actions being more important. At the same time we should note that while “wait" was coded the most, this category represents far from an absolute majority. These data suggest that if people enter a store to buy a present, or have to go to the bathroom, conflict could arise within the group. An explanation can be found in Figure 2: when a portion of the group has to use the bathroom, about 30-40% of the participants indicated the robot should wait. However, when the situation applied to the majority, this number almost doubled to 60%.

*The needs of the many outweigh the needs of the few*

In a similar fashion, when conflict arises it would seem that the needs of the many might outweigh the needs of the few. Figure 1 shows that in three situations participants were less likely to have the robot continue when a situation applied to the majority of the group. Interestingly, these appear to be situations of which the expected duration is unpredictable (situations #2, #9 and #10).

## Discussion & conclusion

In this extended abstract we provided partial results of an explorative study in which we investigated how people's beliefs of robot actions at airport situations. The strength of this study lies in the fact that by collecting free-text responses, participants' were not limited by a pre-defined set of answer categories. To make sense of the data, we coded the data in one of ten categories, on which we applied a majority-voting principle to decide which action a robot should take. A limitation is that this removed the subtleties and richness of the data.

The manipulations were of no significant influence of the coded data, which would suggest that the manipulations do no influence participants' believe of a robot's actions, which could be interpreted as a permit to deploy robots with a single type of response to an action without taking the context into account. At the same time, we have to acknowledge that the limited sample size, caused partly by the resource-intense activity of quantifying qualitative data, is a major reason underlying the lack of significant results. Especially (national) culture has been found of influence to human-robot interaction [3, 11].

With this study, we have made a first step to investigate people's beliefs of robot responses in conflict situations in public spaces, in particular airports. We hope this explorative study benefits (more quantitative) future studies related to this topic intersecting robots, computer science and psychology.

## Acknowledgments

This research was partly supported by the European Commission under contract number FP7-ICT-600877 (SPENCER). The authors would like to thank dr. Manja Lohse for her valuable input on the design of the study.

## REFERENCES

1. Christoph Bartneck, Tatsuya Nomura, Takayuki Kanda, Tomohiro Suzuki, and Kennsuke Kato. 2005. Cultural differences in attitudes towards robots. *Companions: Hard Problems and Open Challenges in Robot-Human Interaction* (2005), 1.
2. Ghadeer Eresha, Markus Haring, Birgit Endrass, Elisabeth Andre, and Mohammad Obaid. 2013. Investigating the influence of culture on proxemic behaviors for humanoid robots. In *22nd IEEE Int. Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 430–435.
3. Vanessa Evers, Nuno Menezes, Luis Merino, Dariu Gavrila, Fernando Nabais, Maja Pantic, Paulo Alvito, and Daphne Karreman. 2014. The development and real-world deployment of FROG, the fun robotic outdoor guide. In *Proceedings of the 2014 ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*. ACM, 100–100.
4. Andy Field. 2009. *Discovering statistics using SPSS (3rd edition)*. Sage Publications Ltd, London, UK. ISBN: 978-1-84787-3.
5. Michiel Joosse, Manja Lohse, and Vanessa Evers. 2015. How a guide robot should behave at an airport-insights based on observing passengers. In *CTIT Technical Report, TR-CTIT-15-01, University of Twente*.
6. Dora C Lau and J Keith Murnighan. 1998. Demographic diversity and faultlines: The compositional dynamics of organizational groups. *Academy of Management Review* 23, 2 (1998), 325–340.
7. Amit Kumar Pandey and Rachid Alami. 2009. A framework for adapting social conventions in a mobile robot motion in human-centered environment. In *Proceedins of the Int. Conf. on Advanced Robotics (ICAR)*. IEEE, 1–8.
8. Rudolph Triebel, KO Arras, Rachid Alami, Lucas Beyer, Stefan Breuers, Raja Chatila, Mohamed Chetouani, and others. 2015. SPENCER: A socially aware service robot for passenger guidance and help in busy airports. In *Proceedings of the 10th Conference on Field and Service Robotics (FSR)*. University of Toronto.
9. Jered H Vroon, Michiel P Joosse, Manja Lohse, Jan Kolkmeier, Jaebok Kim, Khiet Truong, Gwenn Englebienne, Dirk K J Heylen, and Vanessa Evers. 2015. Dynamics of social positioning patterns in group-robot interactions. In *24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 394–399.
10. Michael L Walters, Kerstin Dautenhahn, Sarah N Woods, and Kheng Lee Koay. 2007. Robotic etiquette: results from user studies involving a fetch and carry task. In *2nd ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*. IEEE, 317–324.
11. Lin Wang, Pei-Luen Patrick Rau, Vanessa Evers, Benjamin Krisper Robinson, and Pamela Hinds. 2010. When in Rome: the role of culture & context in adherence to robot recommendations. In *Proceedings of the 5th ACM/IEEE Int. Conf. on Human-Robot Interaction (HRI)*. IEEE Press, 359–366.