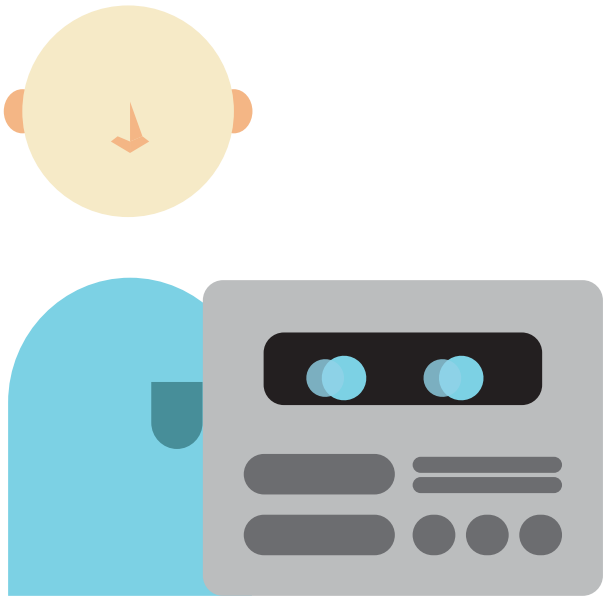
Mutual Understanding for Human-Machine **Collaboration**

Evolving Leader-Follower Behavior

Emma M. van Zoelen

Final Master Project
February 2019 - January 2020
Constructive Design Research track

Mentor: Matthias (G. W. M.) Rauterberg



General **Introduction**

Welcome to the final report for my Final Master Project. This document presents the work I have done over the last year. It was a journey that started with a big question:

How can humans and machines achieve mutual understanding when they collaborate?

I have explored different facets of this question, though always with a focus on emerging and evolving interactions. Eventually mutual adaptation was added on top of that, a mechanism which I believe is vital to achieve smooth collaboration between any combination of intelligent systems.
The final topic which I researched with an experiment is the adaptation and evolution of leading and following behavior by collaborating humans and machines. By observing how participants change their leading behavior over time and by critically looking at when they reconsider their own leading or following role, I have created an overview of how different people deal with seemingly conflicting intentions between themselves and a collaborating machine. I hope that this overview will enable designers and develops of collaborative machines to create positive human-machine collaborations, in which the human feels in control just enough to give the machine the freedom to use its autonomy.
The document in front of you consists of two main parts. The first part you will encounter is an academic paper describing the related work, methods and results of the empirical work that I have done during my final semester at Industrial Design. Secondly, the different parts of the process of the whole year are described. In this second part the focus is on reflecting on the choices I have made and how the different parts connect to each other.
I hope that reading this document will give you interesting new insights in the complex ways in which humans adapt their behavior while interaction with their environment, and more specifically intelligent machines.

- Emma

Table of **Contents**

Paper **Experiment**

The following pages contain the paper I wrote about the experimental work I did this semester. A deliberate decision was made to write this paper from a first person perspective, since this is my graduation work and it feels like using ‘we’ does not do justice to the purpose of this document.

Evolving Leader-Follower Behavior in Human-Machine Collaboration

Emma M. van Zoelen
Department of Industrial Design
Eindhoven University of Technology
Eindhoven, The Netherlands
emmavanzoelen@gmail.com

ABSTRACT
As developments in artificial intelligence progress, more tasks arise in which humans and machines need to collaborate. When their capabilities are complementary, leadership roles will constantly shift. Over time, these shifts become more natural as both systems start understanding their partner's behavior in context of the task. This research explores what such leadership shifts look like, how people adapt their behavior to accommodate these dynamics and how this influences trust and understanding. It was found that explicit and implicit feedback factors from the task and the machine partner trigger humans to reconsider leadership behavior. The development of the extent to which they lead, however, is diverse across participants. Participants who lead more subjectively evaluate the collaboration with the machine lower, but overall the subjective evaluation grows over time. These results support future design of positive collaborative experiences between human and machine, where control can be mediated in a responsible way.

Author Keywords
Human-machine collaboration; human-AI interaction; leader-follower dynamics; human-robot interaction.

CCS Concepts
• Human-centered computing ~ Human computer interaction (HCI) ~ Empirical studies in HCI

• Human-centered computing ~ Human computer interaction (HCI) ~ Interaction paradigms ~ Collaborative interaction

INTRODUCTION
As intelligent and adaptive systems gain more autonomy, they take over more responsibilities of people. As these systems become more ubiquitous, more and more tasks will arise in which humans and machines have to work together to make optimal use of the qualities of both. This has been studied in the academic world for quite a while, especially in safety-critical contexts such as the military [22], space applications [3, 8] and search and rescue problems [15, 16, 21]. More recently, the question whether machines should be considered partners rather than merely tools has arisen as a more general topic in the human-computer interaction and design community too [2]. Eventually, we might say that all interactions between humans and intelligent and adaptive systems are in some way collaborative activities.

When the capabilities of collaborating humans and machines are complementary, situations might arise in which the intentions of the human and the machine are seemingly conflicting, because the partners do not fully understand the mental processes happening in the mind of the other. Such situations cause constant shifts in leadership roles, creating a necessity to balance leading and following by adapting to each other in the collaborative process. After collaborating for a while, these shifts might become more natural as both systems start understanding the behavior of the other to a greater extent in context of the task. Following this, the main purpose of the work presented in this paper is to explore and observe how humans adapt their behavior to deal with such shifting leadership roles and conflicting intentions in human-machine collaborative tasks.

Preliminary experiments inspired by work on language evolution (e.g. [27]) and joint action coordination (e.g. [29]) showed that balancing leading and following is an important mechanism that enables coordination as well as the emergence of interaction patterns that increase the human's understanding of the machine [18]. It can be said that such a process helps to establish and maintain common ground, one of the main aspects necessary for enabling collaboration between humans and machines [13]. This might also be called mutual understanding, meaning that both parties are able to predict and/or explain the other's actions, leading to trust and eventually smooth collaboration [1]. It has also been suggested in literature that both parties need to adapt their behavior to the other to achieve this [26]. Consequently, in the research presented here I attempted to understand how the process of adapting behavior in shifting leadership situations influences understanding of and trust in the machine.

In order to achieve the described objectives, I conducted an experiment in which participants performed a task collaboratively with a Wizard of Oz (WoZ) machine. The task has been defined such that the participants were constantly confronted with seemingly conflicting intentions of the machine. At every moment in the task, the participants had to decide to lead or follow the machine, allowing me to observe when they switched between leading and following and how their behavior changed over time. A detailed description of the experimental methods is given in Section 3. Results consist of the different ways in which leadership

behavior develops over time and how this relates to subjective collaboration fluency, as well as specific triggers for switching between leading and following roles. The results are described extensively in Section 4.

This work builds towards future design of humane and positive collaborative experiences between human and machine, where control can be mediated in a responsible way. It contributes to the recently introduced scientific discipline of 'Machine Behavior', a discipline which includes the study of the interplay between humans and machines at different levels in a behavioral manner [23].

RELATED WORK

Human Social Coordination

Several studies exist on how humans collaborate, that might be used as a basis for understanding human-machine collaboration in which leader/follower roles are coordinated and developed over time. In this section I will focus on coordination in joint action performance related to leader/follower roles.

The research area on coordination in joint action generally focuses on the learnt coordinative behavior. Several studies exist that have looked into 'action signaling' as a way of implicit communication about intentions and (leadership) roles in collaborative activity (e.g. [28]). Some papers evaluate the coordination behaviors when there is an asymmetry in the coordinative task (e.g. predefined leaders and followers) (e.g. [24, 29]). It is shown that by slightly adapting action execution, both leaders and followers signal their role to the other, thereby moderating how the other should coordinate with them. This also happens when roles are not explicitly defined, but determined by the abilities of the two actors. The existing work, however, deals only with situations in which the (leader and follower) roles are static and do not shift or change over time. There is some work on ongoing behavioral dynamics of coordination [14] which shows that humans engage in a mutually adaptive process when they need to coordinate. However, the task used is a very static one and by design the capabilities of the coordinating partners are symmetrical. My research focuses on ongoing behavioral dynamics of coordination, but in tasks where the capabilities of the partners are asymmetrical with shifting asymmetries.

Creating Human-Machine Common Ground

There is some work on creating common ground between humans and machines, generally with a focus on verbal communication. Experimental work generally looks at having a human teach a robot the right words for certain objects [12, 30], and usually the technical problems are emphasized more than any interaction problems. In [12] some interaction problems are mentioned, such as the problem of attention; how do humans know that the robot attends to the right object, considering its own goals and way of thinking? This already touches on a problem that arises when trying to create mutual understanding. Work done by Chai et al. [5] focuses more directly on the human understanding of the machine, as they evaluate when humans believe that common ground is established, and how this can be extended to mutual understanding. Apart from the actual communication behavior, they look at the influence of the general collaborative effort of the robot. It shows that the actions of an agent play a large communicative role in establishing mutual understanding as well, apart from direct communication. This indicates that the general coordinative behavior reaches further than direct communication, and optimizing this behavior to enable smooth collaboration should be situated in the actual collaborative activity rather than merely in a language grounding process. In my work I address the role that behavior has in communicating specifically shifts in leader and follower roles.

Mutual Adaptation

In human-robot, human-computer and human-machine interaction, a lot of work has been done on letting machines (or computers, robots, agents) adapt to humans, also in collaborative activity to enable smooth collaborations (e.g. [6, 12]). However, in those works generally no attention is paid to the fact that the human will adapt their behavior as well, and what such a mutually adaptive relationship might mean for the collaboration. Works that do look at human adaptation either do not pay attention to the influence of the behavior of the intelligent machine (e.g. [17]), or give the intelligent machine the ability to deal with an adaptive human but do not analyze the changes in behavior of the human (e.g. [19]). The purpose of my work is to focus on human adaptation, while taking the behavior of the machine into account. I want to analyze what patterns of behavior emerge and whether those help the human to better understand the intelligent machine's behavior. Ultimately, this can help in creating intelligent machines that can adapt to the human keeping their adaptation into account as well.

Meaningful Human Control

The work presented in this paper relates to a relatively new research area on 'Meaningful Human Control', which has arisen as a response to the future possibility of autonomous weapon systems [25], but is currently considered relevant in many other human-machine interaction or collaboration contexts. In this research area, the main objective is to find a way to make sure that humans have the right amount of control over the behavior of an autonomous system, to make it behave as we humans desire without compromising the benefits of its autonomy. The question of who leads and who follows at which point of the task execution can be translated into a question of who is in control as well.

A lot of the work on Meaningful Human Control (MHC) currently deals with the discussion of what that term actually means, or what the requirements for MHC are. It is for example unclear whether it means that a human should constantly be 'in the loop', or whether humans can also be allowed 'off the loop' [7]. Some papers take a slightly more practical approach by applying philosophical theory to

specific applications, such as autonomous vehicles [4] or surgical robots [10]. However, an (empirical) analysis of what MHC actually entails for humans collaborating with intelligent machines, how their behavior is influenced by the level of control and how it might develop over time is not yet present in these works. While the work presented in this paper does not give methods for achieving Meaningful Human Control, as most of the existing work attempts to do, it does provide an analysis of how leading behavior and as a consequence objective and subjective control develops and evolves over time in a human-machine collaboration.

METHOD

The following section explains the designed setup and interaction that enabled me to evaluate the emergence and evolution of leadership shifts in human-machine collaboration. The described experiment and data collection was reviewed and approved by the Ethical Review Board of the Industrial Design department at Eindhoven University of Technology.

Designing an Interaction

To be able to study evolving leadership shifts between human and intelligent machine, an interaction with such an intelligent machine had to be designed. Taking inspiration from other interactions between humans and non-humans, in this case between blind people and their guide dogs, I designed a remote controlled robot with a leash (Figure 1) and a navigation task. The form of the robot was kept ambiguous and without anthropomorphic features on purpose, to allow anyone interacting with it to focus on the interaction and not the form.

The leash was designed to be the only direct communication channel between the robot and a participant, to ensure specific evaluation of the interaction through the leash without too much noise of other interaction modalities. On top of that, the leash interaction allows for subtle and implicit interactions as both the participant and the robot can pull the leash more or less. The robot was explicitly made to be quite large and heavy, to allow it to pull the participant in a direction as well.

Figure 1. The designed robot used in the experiment.

Figure 2. The field on which the task was executed. Participants moved from the goal on the left (where the robot is stationed) to the goal on the right.

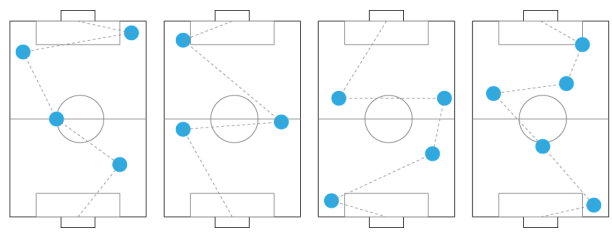


Figure 3. The four predefined maps with the locations of the objects (blue circles), including a line indicating the default route of the robot. The bottom of the field is the starting point.

The fact that the robot was remote controlled enabled the design of a WoZ (Wizard of OZ) experiment, which ensured that the behavior of the robot was relatively controlled but still flexible enough to adapt to the behavior of different participants.

Experimentation Method

Participants were told that they had to perform a collaborative task together with an intelligent robot, while holding the leash of the robot. They were presented with a small football field (Figure 2) and the described WoZ robot, and were told that the aim of the task was to move from one of the goals to the goal on the other side together with the intelligent robot, while scoring as many points as possible. They were given 60 points at the start of the experiment, but lost a point for every second it took them to reach the final target location. They could get extra points by picking up virtual objects that were hidden in the environment, but only the robot knew where those objects were. A sound would indicate that they had picked up an object. This setup required the participants to make a choice between speed or exploration (and thus between leading and following) in order to end up with the highest score. The robot would follow a default route to pick up all virtual objects, unless being pulled away from this route by the human participant.

Before the start of the experiment, the participants were given the possibility to walk from one end of the field to the other with the robot, to give them an indication of the speed of the robot. After that, the first round of performing the task started. The task was performed four times per participant,

where the locations of the virtual objects were different for every round. Four maps with specified locations of the virtual objects were predefined for the Wizard (Figure 3), but the order of these maps was randomized for each participant to make sure that the observed behavior would not be influenced by the specific maps.

Participants

A total of 18 people participated in the experiment (9 male, 9 female), consisting of students from different programs within Eindhoven University of Technology, with an average age of 23 (SD = 3.9). All participants were told that the person with the highest number of points on a single run would receive a gift voucher of €10 to motivate them to perform to the best of their abilities. Before the start of the experiment, the participants gave their consent after carefully reading the consent form that explained all details of the experiment except for the focus of the research (evolving leadership shifts) and the specific behavior of the robot. After the experiment, they were debriefed on the exact purpose of the experiment.

Data Collection

Several types of qualitative and quantitative data were collected. First of all, before starting the first round of the experiment, participants completed a short Big-5 personality questionnaire [9]. While performing the task, a camera placed in one of the corners of the field recorded the behavior of the participants. Furthermore, after each round of the task participants were asked to complete a questionnaire on subjective Human-Robot Collaboration Fluency [11], as well as to answer the following three interview questions:

1. Can you explain the behavior of the robot?
2. What was your strategy for completing the task?
3. How did you experience the collaboration?

Data Analysis

For one of the female participants data was missing in the Collaboration Fluency questionnaire. Therefore, for the analyses relating this measure this person was excluded.

A quantitative analysis of the Collaboration Fluency questionnaire to see if the scores changed over the different rounds was done using a Repeated Measure ANOVA. A post-hoc Tukey HSD test with Bonferroni correction was used to test between which rounds any found differences were significant.

Furthermore, in order to enable comparison between behavioral, qualitative data and data from the Personality questionnaire, an exploratory k-means cluster analysis was performed on the Personality questionnaire data. In this analysis, a participant was taken as one object to be clustered (meaning that an object consisted of five variables for the different personality factors). This allowed me to evaluate clusters of participants with similar responses and later compare those to clusters of participants with similar behaviors.

Code Category	Code	L/F/N
Task events	Task is running	N
	Object Sound	N
Robot movement	Standing still	N
	Moving towards object	L
	Moving away from goal	L
	Moving across field	L
	Moving with participant	F
	Moving in goal direction	F
Participant movement	Standing still/waiting	F
	Moving around robot	N
	Moving in goal direction	L
	Moving in robot direction	F
	Moving across field	L
Leash activity	Loose	F
	Stretched	L
	Pulled in direction	L
	Loosening/Stretching	N
Participant location relative to robot	Behind	F
	In front of	L
	Next to	N

Table 1. The closed coding scheme, including the code categories, the exact codes used and the characterization of leading (L), following (F) or neutral (N) behavior.

The videos were coded using an open coding process at first, to get a view on the different kinds of behavior present among participants as well as on events that triggered participants to switch between a more leading and a more following role. From the open codes, a coding scheme for closed coding was developed that contained codes for task events, robot movement, participant movement, leash activity and the participant's location relative to the robot. Each code was characterized as a leading, following or neutral behavior (Table 1).

All videos were then coded again using a closed coding process using The Observer XT [20], enabling me to visually analyze the different behaviors across rounds simultaneously (Figure 4) as well as to quantify the amount of leading behaviors present in each round. Intercoder reliability for the durations of sequences with another coder for 5.6% of the data (4 videos) was found to be 97.55%.

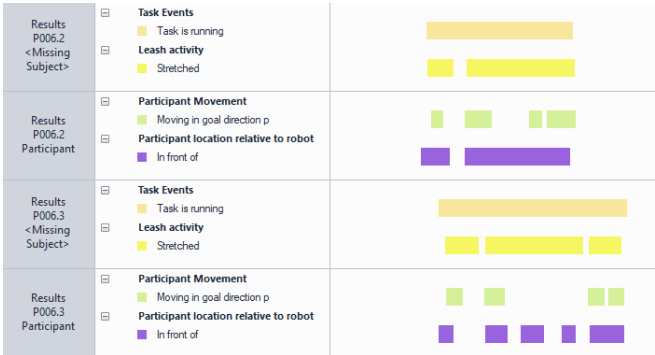


Figure 4. A screenshot from the visualization made in The Observer. These are the duration of the task and the different leading behaviors for participant 6, round 2 and 3.

Looking at the development of leading behavior across categories of participant behavior (leash activity, participant movement and participant location), participants with similar behavior were clustered. In a similar manner, participants were again clustered based on interview answers.

In the final part of the analysis, I compared results of different measures in an attempt to explain some of the observed behavior. Both the interview clusters as well as the personality clusters were compared to the behavioral clusters visually using crosstabulation. Finally, the leading behaviors across the categories of participant behavior were tested on correlation with the scores on the Collaboration Fluency questionnaire for each round using a Pearson correlation test. In order to do this, for each leading behavior the total duration relative to the total duration of the task was taken per participant per round.

RESULTS

Collaboration Fluency

The Repeated Measure ANOVA of the results of the Collaboration Fluency questionnaire for all participants indicated a significant difference between the different rounds ($F(3,48) = 6.76$, $p < .001$). The post-hoc analysis showed significant results between rounds 1 ($M = 51.11$, $SD = 9.80$) and 3 ($M = 58.86$, $SD = 10.23$), between rounds 2 ($M = 54.49$, $SD = 9.19$) and 4 ($M = 62.39$, $SD = 12.97$) and between rounds 1 and 4 (see Table 2). No significant differences were found between male and female participants.

From these results, it is clear that the subjective Collaboration Fluency grows over time, and that this growth can be seen within three rounds. This means that regardless of how people behave, the fact that they interact with the robot by itself makes most people more positive about the collaboration and makes them trust the robot more.

Personality Types

The exploratory k-means cluster analysis showed that about six clusters of different personalities can be made from the outcomes of the Big-5 Personality questionnaire (see Table 3).

Rounds Compared	P-value
1 – 2	1.000
1 – 3	0.023*
1 – 4	< 0.001**
2 – 3	0.623
2 – 4	0.020*
3 – 4	1.000

Table 2. The p-values for the different compared rounds after the Tukey HSD test. Values with * are significant, values with ** are extremely significant.

Looking at the scores for the clusters qualitatively, it is clear that overall the participants do not differ much in personality. On the Agreeableness factor, the Neuroticism factor and the Intellect/Imagination (or Openness) factor almost all clusters have a similar score. This is probably due to the fact that all participants were students at the same university and therefore a quite homogeneous group.

Switch Triggers

The open coding process revealed six types of moments in the task that typically trigger participants to rethink whether they should behave in a more leading or following way. This could be observed by seeing a clear switch, hesitating behavior or several short switches after one of the described triggers.

The first one is quite trivial, namely at the start of the task; participants have an idea of whether they want to start the task in a more leading or following manner prior to starting. The other five types are the following:

	Participants in Cluster	E	A	C	N	I
1	9, 13, 14, 18	4.00 (0.35)	4.94 (0.13)	1.94 (0.43)	3.12 (0.31)	4.75 (0.20)
2	3, 11, 15	3.58 (0.63)	4.33 (0.14)	2.50 (0.25)	2.25 (0.25)	4.00 (0.43)
3	8, 12, 17	3.92 (0.38)	4.08 (0.14)	3.50 (0.25)	2.92 (0.28)	4.50 (0.43)
4	4, 10	2.63 (0.53)	3.13 (0.53)	3.13 (0.53)	3.25 (0.35)	4.25 (0.00)
5	2, 7	3.13 (0.18)	4.25 (0.00)	4.13 (0.18)	2.50 (0.00)	3.50 (0.00)
6	1, 5, 6, 16	2.88 (0.14)	4.38 (0.43)	2.94 (0.90)	3.81 (0.38)	4.56 (0.24)

Table 3. The clusters of participants for the scores on the Personality questionnaire. The different scores stand for Extraversion (E), Agreeableness (A), Conscientiousness (C), Neuroticism (N) and Intellect/Imagination/Openness (I). The values are the mean scores in the cluster, with the standard deviation between brackets.

- Object sound;
- Leash pull by the robot (for example when it goes in another direction than the participant/than expected by the participant);
- The robot deviating from the route that leads to the final goal;
- Getting close to the goal;
- The robot standing still.

These five types can be categorized in explicit, implicit or ambiguous feedback, as well as in task or partner feedback as can be seen in Table 4. Related to the 'partner feedback' category, it was evident that almost all participants implicitly follow the robot when it is going roughly in the direction of the goal. This is clear from the fact that participants often do not move to the goal in a straight line, but follow the robot to move via a side of the field (and thus via one of the virtual objects). The implicit partner feedback is therefore only really a switch trigger if the deviation from the route is substantial. The implicit task feedback was evident as for most participants, being further along in the task made them change their behavior. While this is in essence a gradual change, it was sometimes observable as an immediate change right after people crossed the middle of the field.

Leading Behavior Development

After closed coding, it was possible to look at the development of certain behaviors over the different rounds per participant. I qualitatively looked at the three main categories of behaviors (leash pull, position relative to robot, movement, see Figure 4) and created descriptions for the development of the behavior in each category. I then grouped participants with similar developments into a few main groups (Table 5, Table 6, Table 7). Each behavioral development category was characterized as either a leading, following or balancing behavior.

It is clear that for all three different leading behaviors, similar behavioral developments exist. Also, many participants appear in similar clusters across the different kinds of behavior (e.g. a generally leading behavior for leash activity usually means that there is also a generally leading behavior in the other two categories, or at least a balancing behavior development). In the visualizations of the closed codes this was visible as quite often (though not always) different leading behaviors appeared simultaneously. An interesting difference between the three kinds of behaviors is that for leash activity, most participants start with a stretched leash. A part of those then continue to balance this out with following behavior, but several others do not show development in this behavior. For relative position, however, most participants remain in a leading position for most of the experiment. Last, the movement of the participants is generally more balancing in the sense that most participants level out their behavior over the rounds.

	Explicit	**Implicit**	**Ambiguous**
Task Feedback	Object sound	Getting close to the goal	*Robot standing still*
Partner Feedback	Leash pull by robot	Robot deviating from goal direction	Robot standing still

Table 4. The different types of feedback that trigger reconsideration of leading behavior categorized.

Interview Insights

The interviews were mostly meant to get a view on how participants would subjectively describe the collaboration with the robot in relation to the development of their behavior. For that reason, a description of the development of their answers was made, after which these descriptions were clustered as well (Table 8). For almost all participants it was the case that they started understanding the behavior of the robot better over the course of the rounds, since their explanations of the robot behavior as well as their own strategies became more elaborate and less uncertain. For the great majority of the participants, they also felt like the collaboration became better and more balanced over time (which was also clear from the collaboration fluency questionnaire). Several participants, however, clearly expressed the fact that the collaboration remained imbalanced until the last round, because one of the partners was clearly more important in achieving the task. While several participants regarded this as a negative aspect, there were also some who did not consider this to be a problem or even consciously expressed that they considered this a positive aspect.

Leash behavior development	**Participants showing this behavior**	
1. Start very loose, becomes more stretched over rounds	1, 14	B
2. Start very stretched, becomes more loose over rounds	3, 7, 11, 15, 16, 18	B
3. Loose mostly	4	F
4. Stretched mostly	2, 6, 8, 9, 10, 17	L
5. Stretched in the beginning, loose in the middle, stretched at the end	5, 12	L
6. Loose in the beginning, stretched in the middle, loose at the end	13	F

Table 5. The clusters for leash behavior development, including categorization into balanced (B), following (F) or leading (L) behavior.

Relative position behavior development	Participants showing this behavior	
1. Start next to or behind robot, becomes more in front of robot over time	7, 10, 14	B
2. Start in front of robot, becomes more next to or behind robot over time	3, 18	B
3. Mostly next to or behind robot	4, 15	F
4. Mostly in front of robot	2, 8, 9, 11, 17	L
5. Start in front of robot, next to or behind robot in the middle, in front of robot at the end	1, 5, 12, 16	L
6. Start next to or behind of robot, in front of robot in the middle, next to or behind robot at the end	13	F
7. Start steadily in front of robot, but the behavior becomes more fragmented over time	6	L

Table 6. The clusters for relative position behavior development, including categorization into balanced (B), following (F) or leading (L) behavior.

Movement behavior development	Participants showing this behavior	
1. Start very following, increase of leading movement over time	1, 2, 10, 12, 14	B
2. Start very leading, increase of following movement over time	3, 4, 6, 7, 8, 16	B
3. Mostly following movement	11, 15	F
4. Mostly leading movement	9, 17	L
5. Start very leading, following in the middle, leading at the end	5	L
6. Start very following, leading in the middle, following at the end	13,18	F

Table 7. The clusters for movement behavior development, including categorization into balanced (B), following (F) or leading (L) behavior.

Subjective description of collaboration	Participants in this cluster
Collaboration becomes better and more balanced	1, 2, 6, 7, 8, 10, 11, 14, 16, 17
Collaboration remains imbalanced, this is regarded negatively	3, 12, 13, 18
Collaboration remains imbalanced, this is regarded positively	4, 9, 15
Robot is more of a tool that needs to be lead	5

Table 8. The clusters for the interviews

It is interesting to note that the extent to which the collaboration is balanced was expressed so clearly. Overall, participants considered a balanced collaboration in which both partners contributed superior to an imbalanced collaboration, even if the latter made it easier to complete the task. Some participants even described how they did their best to find aspects in which they could help the robot with its weaknesses, such as the fact that it is quite slow (participants pull the leash and use verbal encouragement to motivate it to be faster) or the fact that it cannot make sharp turns (one participant picked up the robot and carried it the final distance to the goal because they knew it would take the robot a lot of time to make the turn).

Interviews linked to Behavior

Using crosstabulation, it can be seen that most participants that report in the interview that the collaboration becomes better and more balanced also have a movement development that becomes more balanced (8/10). For their leash behavior, they either have a balancing development or have a generally leading behavior (5/10, 5/10), which is similar for their position development (3/10, 7/10). For both imbalanced interview categories the distribution over behavior development styles is quite equal. For the one participant that considered the robot to be more of a tool that needed to be lead, all behavioral measures fall within the leading category.

These numbers suggest that exploring different leading and following behaviors (thus balancing them out) helps participants to empathize with the robot. Also, it seems that participants project their own behavior onto the robot; if they themselves behave in a more balancing way, they consider the robot and general task execution to be more balanced and cooperative.

Collaboration Fluency linked to Behavior

Correlating the total duration of the behavior corrected for the time they took to finish the task to the scores on the subjective Collaboration Fluency yielded the correlation coefficients presented in Table 9.

	Correlation coefficient	P-value
Stretched leash	-0.344	0.006**
Position in front of robot	-0.113	0.373
Moving in goal direction	-0.290	0.020*

Table 9. The results for the correlation test between the total duration of leading behavior and the subjective Correlation Fluency score. Values with * are significant, values with ** are extremely significant.

There is a weak negative correlation between the total duration of a stretched leash and the subjective Collaboration Fluency score. This means that when participants stretch the leash more, they will generally score lower on the questionnaire. A similar but slightly weaker negative correlation can be found between the total duration of the participant moving in the direction of the goal and the subjective Collaboration Fluency score. This suggests that if people portray less (explicit) leading behavior, the Collaboration Fluency is higher. Following from this, when people are less willing to follow the robot, they regard the robot as less cooperative as well. This confirms what was found in the combination of the interviews and the observations: participants project their own behavior onto the robot (if they lead more they consider the robot to be less cooperative), and mixing leading and following helps to empathize with the robot.

Personality Types linked to Behavior

Crosstabulation was used to visually determine whether the behavior of participants was influenced by their personality. The only clear trend that could be seen was that people with leading behavior are mostly in personality category 3 (leash: 3/8, position: 3/10, movement: 1/3) or category 6 (leash: 2/8, position: 4/8, movement: 1/3). Since we determined above that Agreeableness, Neuroticism and Imagination were similar for all participants, we can say that those participants either had medium or medium high Extraversion *and* Conscientiousness. In the other personality clusters, these two scores were more dissimilar from each other. However, since all personalities were relatively similar according to the questionnaire, not much can be said about this.

DISCUSSION

The research presented in this paper explores how leading behavior of humans develops over time when they collaborate with a machine partner, focusing on what triggers them to shift leadership from or to the machine. Looking at leading and following in terms of ongoing behavioral coordination in which the capabilities of both partners are asymmetrical has not been thoroughly studied before. I therefore provide a new perspective on what such collaborative and coordinative behaviors could look like, while explicitly taking into account that human behavior will be continuously influenced by machine behavior in such situations. The results presented serve as an exploration of the different relevant aspects that play a role in these collaborative interactions.

The fact that subjective Collaboration Fluency grows significantly the longer people collaborate with the machine regardless of their specific behavioral development has not been reported before. This result might imply that humans are inclined to trust machines more the longer they collaborate with them, and we might wonder whether that is desirable. Especially in safety-critical environments, it will be important to ensure that this mechanism does not create overtrust. An interesting direction for future work will therefore be to evaluate to what extent robot behavior influences this growth of subjective Collaboration Fluency, and if it is possible to see when the growth slows down or stops.

Related to the above result, the finding that more following behavior results in a higher subjective Collaboration Fluency score contributes a possible risk to the 'meaningfulness' of Meaningful Human Control. If more following people are more happy with their human-machine collaboration, we need to find a way to ensure that people still lead when this is necessary. Again, further research is necessary to determine the extent of this effect and to possibly find design options that ensure that the human remains critical to the behavior of the machine.

The different types of explicit and implicit feedback from partner and task that were uncovered might be of help in ensuring a balanced leader-follower relationship between human and machine. Using triggers for a leadership shift, or at least a leadership reconsideration, could prove useful to enforce a human partner to reevaluate whether they should maybe take over control or take the lead from a machine partner.

Regardless of the triggers for leadership reconsideration, however, the way in which leading and following behavior develops over time differs greatly per person. This might be because some people simply take more time to balance out leading and following, but it can just as well be the case that some people just prefer to be more or less in control. Related to this, depending on the task at hand we might wonder whether it is desirable if all people portray a similar balancing out behavioral development. The diversity of leadership development will need to be taken into account while designing collaborative interaction between humans and machines, and depending on the context at hand a choice between compensating for or encouraging specific behavior must be made.

While the insights gained from the presented experiment provide interesting insights, it is difficult to draw clear conclusions from combining the different measures due to the wide variety of results and the relatively small amount of participants. Further evaluations of the different results are necessary to get a deeper insight. Therefore, the presented

study mostly serves as a first step into understanding the different interaction mechanisms that build up a shifting leader-follower relationship between humans and machines in collaborative activity.

CONCLUSION

Understanding how leading and following behavior evolves in human-machine collaborations requires experimental work that allows for subtle and implicit interactions. Implicit feedback extracted from the behavior of the machine partner as well as the progression of the task triggers participants to reconsider their current leading or following role, allowing for dynamic shifts of leadership throughout the task. Subjective Collaboration Fluency is higher when people are more following, as following allows them to empathize with the machine, but overall this subjective measure grows over time. Apart from these conclusions, however, it is important to note that a wide variety of behavioral developments exists between participants, which must be taken into account in future design of collaborative machines. The discussed results can serve as pointers for further research into the dynamics of leading and following behavior in human-machine collaborations, to eventually create practical strategies for creating and maintaining Meaningful Human Control.

REFERENCES

[1] Azevedo, C.R.B., Raizer, K. and Souza, R. 2017. A vision for human-machine mutual understanding, trust establishment, and collaboration. *2017 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA)* (Savannah, GA, USA, Mar. 2017), 1–3.

[2] Bellamy, R.K.E., Andrist, S., Bickmore, T., Churchill, E.F. and Erickson, T. 2017. Human-Agent Collaboration: Can an Agent be a Partner? *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '17* (Denver, Colorado, USA, 2017), 1289–1294.

[3] Bradshaw, J.M., Sierhuis, M., Acquisti, A., Feltovich, P., Hoffman, R., Jeffers, R., Prescott, D., Suri, N., Uszok, A. and Van Hoof, R. 2003. Adjustable Autonomy and Human-Agent Teamwork in Practice: An Interim Report on Space Applications. *Agent Autonomy*. H. Hexmoor, C. Castelfranchi, and R. Falcone, eds. Kluwer Academic Press. 243–280.

[4] Calvert, S.C., Heikoop, D.D., Mecacci, G. and van Arem, B. 2019. A human centric framework for the analysis of automated driving systems based on meaningful human control. *Theoretical Issues in Ergonomics Science*. (Dec. 2019), 1–29. DOI:https://doi.org/10.1080/1463922X.2019.1697390.

[5] Chai, J.Y., She, L., Fang, R., Ottarson, S., Littley, C., Liu, C. and Hanson, K. 2014. Collaborative Effort towards Common Ground in Situated Human-Robot Dialogue. *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Mar. 2014), 33–40.

[6] Crandall, J.W., Oudah, M., Tennom, Ishowo-Oloko, F., Abdallah, S., Bonnefon, J.-F., Cebrian, M., Shariff, A., Goodrich, M.A. and Rahwan, I. 2018. Cooperating with machines. *Nature Communications*. 9, 1 (Dec. 2018), 233. DOI:https://doi.org/10.1038/s41467-017-02597-8.

[7] Crootof, R. 2016. A Meaningful Floor for "Meaningful Human Control." *Temple International and Comparative Law Journal*. 30, 1 (2016), 53–62.

[8] Diggelen, J. van, Bradshaw, J.M., Grant, T., Johnson, M. and Neerincx, M. 2009. Policy-Based Design of Human-Machine Collaboration in Manned Space Missions. *2009 Third IEEE International Conference on Space Mission Challenges for Information Technology* (Pasadena, CA, USA, Jul. 2009), 376–383.

[9] Donnellan, M.B., Oswald, F.L., Baird, B.M. and Lucas, R.E. 2006. The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five Factors of Personality. *Psychological Assessment*. 18, 2 (2006), 192–203. DOI:https://doi.org/10.1037/1040-3590.18.2.192.

[10] Ficuciello, F., Tamburrini, G., Arezzo, A., Villani, L. and Siciliano, B. 2019. Autonomy in surgical robots and its meaningful human control. *Paladyn, Journal of Behavioral Robotics*. 10, 1 (Jan. 2019), 30–43. DOI:https://doi.org/10.1515/pjbr-2019-0002.

[11] Hoffman, G. 2019. Evaluating Fluency in Human–Robot Collaboration. *IEEE Transactions on Human-Machine Systems*. 49, 3 (Jun. 2019), 209–218. DOI:https://doi.org/10.1109/THMS.2019.2904558.

[12] Kaplan, F. 2000. Talking aibo : First experimentation of verbal interactions with an autonomous four-legged robot. *Learning to Behave: Interacting agents CELETWENTE Workshop on Language Technology* (2000), 57–63.

[13] Klein, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R. and Feltovich, P.J. 2004. Ten challenges for making automation a "team player" in joint human-agent activity. *IEEE Intelligent Systems*. 19, 6 (Nov. 2004), 91–95. DOI:https://doi.org/10.1109/MIS.2004.74.

[14] Konvalinka, I., Vuust, P., Roepstorff, A. and Frith, C.D. 2010. Follow you, Follow me: Continuous Mutual Prediction and Adaptation in Joint Tapping. *Quarterly Journal of Experimental Psychology*. 63, 11 (Nov. 2010), 2220–2230. DOI:https://doi.org/10.1080/17470218.2010.497843.

[15] Kruijff, G.J.M., Janıcˇek, M., Keshavdas, S., Larochelle, B. and Zender, H. 2012. Experience in System Design for Human-Robot Teaming in Urban Search & Rescue. *8th International Conference on Field and Service Robotics* (Matsushima, Japan, 2012), 1–14.

[16] Mioch, T., Peeters, M.M.M. and Neerincx, M.A. 2018. Improving Adaptive Human-Robot Cooperation through Work Agreements. *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Aug. 2018), 1105–1110.

[17] Mohammad, Y. and Toyoaki Nishida 2008. Human adaptation to a miniature robot: Precursors of mutual adaptation. *RO-MAN 2008 - The 17th IEEE International Symposium on Robot and Human Interactive Communication* (Aug. 2008), 124–129.

[18] Mutual Understanding for Human-Machine Collaboration: 2019. *www.emvanzoelen.com/files/M2.1_mutualunderstanding.pdf*. Accessed: 2020-01-08.

[19] Nikolaidis, S., Hsu, D. and Srinivasa, S.S. 2017. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *I. J. Robotics Res.* 36, (2017), 618–634. DOI:https://doi.org/10.1177/0278364917690593.

[20] Noldus, L.P.J.J. 2019. *The Observer XT*.

[21] Nourbakhsh, I.R., Sycara, K., Koes, M., Yong, M., Lewis, M. and Burion, S. 2005. Human-Robot Teaming for Search and Rescue. *IEEE Pervasive Computing*. 4, 1 (Jan. 2005), 72–78. DOI:https://doi.org/10.1109/MPRV.2005.13.

[22] Parasuraman, R., Barnes, M. and Cosenzo, K. Adaptive Automation for Human-Robot Teaming in Future Command and Control Systems. *The International C2 Journal*. 1, Decision Support for Network-Centric Command and Control, 43–68.

[23] Rahwan, I. et al. 2019. Machine behaviour. *Nature*. 568 (2019), 477–486.

[24] Sacheli, L.M., Tidoni, E., Pavone, E.F., Aglioti, S.M. and Candidi, M. 2013. Kinematics fingerprints of leader and follower role-taking during cooperative joint actions | SpringerLink. *Experimental Brain Research*. 226, 4 (2013), 473–486.

[25] Santoni de Sio, F. and van den Hoven, J. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. *Frontiers in Robotics and AI*. 5, 15 (2018), 1–14. DOI:https://doi.org/10.3389/frobt.2018.00015.

[26] Sciutti, A., Mara, M., Tagliasco, V. and Sandini, G. 2018. Humanizing Human-Robot Interaction: On the Importance of Mutual Understanding. *IEEE Technology and Society Magazine*. 37, 1 (Mar. 2018), 22–29. DOI:https://doi.org/10.1109/MTS.2018.2795095.

[27] Scott-Phillips, T.C., Kirby, S. and Ritchie, G.R.S. 2009. Signalling signalhood and the emergence of communication. *Cognition*. 113, 2 (2009), 226–233.

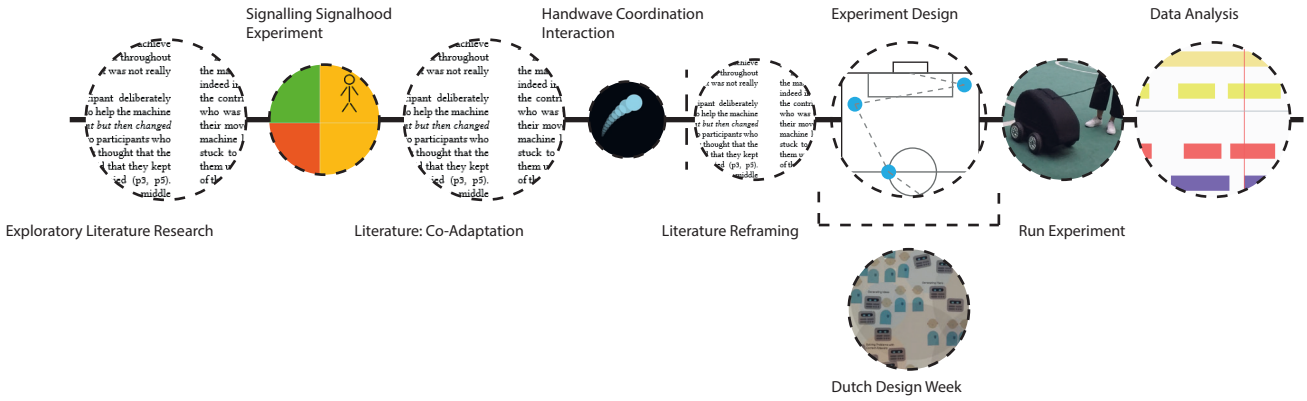
[28] Vesper, C. 2014. How to support action prediction: Evidence from human coordination tasks. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh, UK, Aug. 2014), 655–659.

[29] Vesper, C. and Richardson, M.J. 2014. Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research*. 232, 9 (Sep. 2014), 2945–2956. DOI:https://doi.org/10.1007/s00221-014-3982-1.

[30] Zhang, H., Yu, H. and Xu, W. 2018. Interactive Language Acquisition with One-shot Visual Concept Learning through a Conversational Game. *arXiv:1805.00462 [cs]*. (Apr. 2018).

FMP **Process**

In the following sections, all steps of my FMP process (both M2.1 and M2.2 semester) will be discussed. The focus will be mostly on why I made certain decisions and the deliberations I had before making them. The different process parts are presented in chronological order.



M2.1 Semester

Exploratory Literature Research

The original motivation for the project presented in this report came from an interest in the evolution of language as a way to understand our world, combined with an interest in how artificial cognition differs from human cognition. Since my general research interest is in social interaction between humans and intelligent machines, I wondered: what if we would let interaction languages emerge between humans and machines? Would that help us to truly understand each other? For a part this is more of a philosophical question than a practical one, but as I knew that work exists on human as well as on artificial language evolution, I thought it would be interesting to explore what such a process could look like. Therefore, my initial literature research explored papers in the areas of human cognition, language evolution, algorithms for social coordination among agents and human-computer communication. The literature review was very broad, and therefore generated large research questions:

1. What kind of intelligent adaptive system enables the emergence of language for mutual understanding between humans and intelligent adaptive systems?

2. How does mutual understanding develop between a person and an intelligent adaptive system?

While these two questions are definitely too broad for the scope of the project that I did, they did provide me with the initial direction and ultimately the larger question that I hope my research contributes to.

Signaling Signalhood Experiment

After reading the literature about social coordination in language evolution, it struck me that most of the existing work looks into the kind of signals that develop between people. There is very little focus on how people manage to coordinate their signals and the underlying interaction mechanisms. Since the focus of my project was to create interactions between humans and intelligent machines that allowed for the coordination of an interaction language, I wanted to get a more in depth understanding of the process of coordinating signals.
In order to achieve this, I decided to stay close to the literature and replicate an existing experiment. I chose the 'Signaling Signalhood' experiment [1], which focuses on language coordination in a situation where the distinction between task actions and communication actions is not predefined. In the original experiment, the researchers again mostly looked into the type of signals that pairs developed, but since I wanted to look at the mental processes that happen in people during the coordination as well as the interaction mechanisms, I decided to let my participants think aloud during the experiment. This, combined with short interviews afterwards and observations of their behavior, helped me to qualitatively understand how people attempt to achieve coordination and the mechanisms that support this.
The main interaction mechanisms observed that were used in the rest of the process were the following:

1. Imitation or mimicry, as imitating the behavior of a partner was used both as a confirmation of a decision as well as a way of validating the meaning of the signal;

2. Balancing leading and following or simply initiative-taking in defining signals, as at all times both participants had to be open to following their partner if they observed that the partner was sticking to a specific signal, as well as being open to the opportunity of initiating new signals themselves when their partner was showing confusing or contradictive behavior.

Literature: Co-adaptation

The findings from my first experimental work lead me to look more into theories about social coordination. I came across several theories and interaction mechanism that described a process in which interacting agents develop patterns of acting together by adapting to each other. The main terms I found were co-evolution, co-adaptation, co-performance, co-learning and as a more general theory social practices. All of these described to a certain extent the interaction mechanism that I had in mind. Having very little experience with developing a theory myself, I had some difficulties in condensing all this literature into a single theoretical background that could be used as a frame of reference in my research. I did however keep the idea of mutual adaptation to achieve social coordination as a focal point throughout the rest of the project.

Handwave Coordination Interaction

I wanted to translate my findings from the first exploratory experiment and the literature reviews into a designed interaction. I stayed quite close to the initial research direction that was based on language, and decided to create an interaction to coordinate a movement sign made by waving your hand around. Moreover, I focused on the two interaction mechanisms identified in my previous experiment, imitation and finding a balance in leading and following.
The designed interaction was very simple and the prototype was not very robust. However, it did express the ideas that I had in a tangible way, helping me to slightly concretize them. Also, it showed me that my research direction was much too vague at that point, causing the insights from a small evaluation with participants to be very abstract and simply confirming the insights from the previous experiment.

The animation of the handwave coordination interaction.

M2.2 Semester

Literature Reframing

After the assessment of the M2.1 semester, it was clear that the scope of my research was still way too broad, which limited the progression of a clear research design. Initially, I did an exploration of several contexts to which my earlier ideas and work could be applied. This was not easy for me, as I kept seeing large gaps between my initial research questions and the questions I could ask in contexts I could come up with. Other ideas still had too broad of a scope. Eventually, I explored two contexts in a bit more detail:

Autonomous Vehicle to Human Road User Communication

A possible context to which the idea of the emergence of interaction languages and mutual adaptation can be applied, is that of communication between human road users and autonomous vehicles. The research on this topic has recently grown, but much of the existing work focuses on set messages shown via text, images or light on or around the car. Considering the fact that communication between human road users is much less implicit and developed for its purpose on the go, as well as the fact that some of these communications might be culture-depended, made me think about the possibilities of having such communication emerge and evolve. While there are of course some limitations to such a system in terms of safety, it can definitely be an interesting idea worth investigating, especially to challenge the current work being done to be a bit more out-of-the-box. Eventually I did not continue this idea, for two main reasons. First of all, the interaction presented above would happen across a population of people, whereas my initial ideas and research questions were focused more on a one-person, one-machine interaction. This direction would therefore make the project too different from what I envisioned. Moreover, it was very difficult for me to design a research setup that allowed me to answer my research questions in a feasible manner and within the time limits.

USAR Scenario

After the autonomous vehicle ideas, I started exploring contexts related to an Urban Search and Rescue task. Since much of the existing research on human-agent teaming has been done within this or related contexts, it seemed like a useful direction to pursue. However, the interaction concepts that I came up with were very focused on having the human tell a machine which objects to collect in a dangerous environment. This meant that the main topic of the research would still be on naming of objects and language, which would be very similar to existing research on grounded communication. Therefore I decided that more literature research was necessary to frame my final research direction.

This time, I focused my literature research on mutual adaptation in human-machine collaboration specifically. I came across several works that I had not seen before because of using different search terms that attempted to research this topic (e.g. [2], [3]). I also found that while these works were very interesting, they generally looked at adaptation in a one-sided manner, focusing very much on behavior of one of the interacting partners. Moreover, they were either still very focused on the adaptation of symbols and signs, or did not qualitatively study the adaptive behavior of the human. Using these new insights, I reframed my research to focus on the qualitative analysis of the adaptation of human behavior to an intelligent machine. In this, leader-follower dynamics would be the core topic, since this was one of the interaction mechanisms I uncovered during my first semester.

To make sure that the final task would actually be a collaborative interaction between human and machine, where the machine would be a partner rather than a tool, I looked into some literature about the definition of a tool versus a partner. A few people have attempted to define the difference between the two, where the distinction described in [4] seemed to be the clearest. In this work it is described that “a tool [is] an entity that is used in order to extend a person’s capabilities and efficiency in carrying out a task” (p. 3). They describe a more advanced category of tools, namely adaptive tools, in which the tool “has the ability to change one or more of its own parameters in response to environmental variations” (p. 4). To extend on that, a cooperative assistant or collaborating partner needs to be able “to model the behavior of another agent in relation to a goal as well as its own actions and abilities” (p. 5). It therefore needs to show behavior that takes into account the effect of its partner’s actions on the goal. I kept this definitions in mind while designing the interaction and task for the experiment.

Experiment Design

Informative Interviews

In the process of defining the experiment and interaction for my research, I interviewed two people. These conversations helped me to get a grip on the kind of context and task situation I was looking for, as well as the subtle interactions necessary for building trust and understanding.

Cordula Vesper

Cordula Vesper is a researcher on intentional joint action coordination. In her work, she focuses mostly on the cognitive mechanisms underlying coordinative behavior, relating this to action and communication as well. Several of her papers describe leader-follower dynamics specifically. I interviewed her to get a better understanding of the existing research in this area as well as the relevant research questions and tasks. Cordula explained that if there is an asymmetry in knowledge, leader/follower roles will automatically appear as the asymmetry pushes agents in their specific role. An interesting situation to look at is what happens when the advantage of the asymmetry switches to another agent many times throughout the task. Leader/follower roles must then be negotiated. This can be done in an implicit or explicit matter, but making it explicit makes the definition of leadership very binary. By looking at more small scale adaptations and making it more subtle (e.g. not 'do you make a detour from your own course of action, yes or no' but 'how much of a detour do you make') it is possible to look at the dynamics of shifting leadership. Of course one of the main questions for the agents is then how they can communicate their knowledge and decisions to the other. Information about this can be found in literature on continuous synchronization (e.g. [5]).

Existing studies on leader/follower dynamics do not study the negotiation and shifting of these roles, and as far as Cordula knows there are no studies about how the signaling of those roles might emerge.

The main message was to keep the task simple, as it is easier to study these topics when the dynamics are more subtle. There should be only one level of leader/follower dynamics, and it is best to focus on this aspect only and not add in imitation as another main topic (although I might allow for it). It can then possibly be measured in different scenarios.

Nicole Pedder (KNGF)

Nicole Pedder is a trainer of guide dogs for blind people. I interviewed her to get insights into how humans and guide dogs learn to collaborate, as this is also an example of a human and a non-human having to coordinate their actions together and understand each other in the context of their task.

From the interview, a few main points emerged that were interesting for or applicable to my context. First of all, Nicole told me that guide dogs always take the lead in walking outside, while the human only determines the route. By default, the human should follow the dog always. However, it is the responsibility of the human to make sure that the dog sticks to trained behavior. The human can decide to teach the dog new behavior, or to change or remove trained behavior, but it is important to remember that dogs are not good at dealing with exceptions. The behavior of the dog can only reach a certain level of detail. Basically, it is the human's responsibility to make sure exceptions are handled in such a way that the dog does not get confused.

For the human, the most difficult part of learning how to work with a guide dog is letting the dog take over control. Partly because of this, in training together, the focus is on building trust and building the general bond between human and dog. A vital quote from our conversation that followed the above insights, was the following: "There is a necessity for an ability to adapt on both sides to build trust."

Task definition

In order to define the task that would allow me to evaluate leader-follower dynamics, I set a few requirements on top of the insights from the interviews. These requirements were based on the work I did during my M2.1 semester and insights gained from literature research:

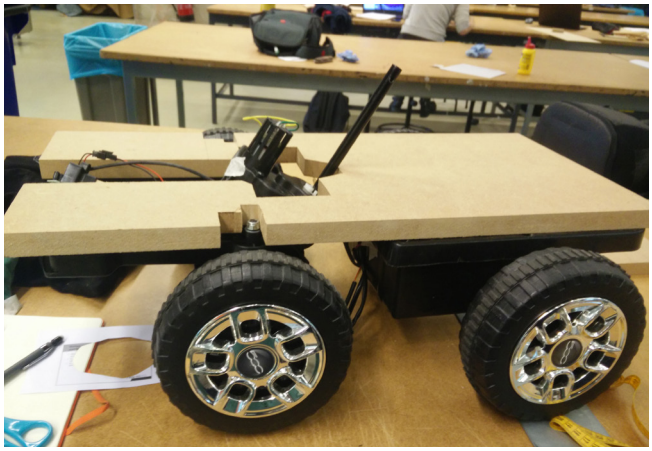
- Both parties can take initiative (machine is active, this is not the case in the work of [2]. It is in the work [3], but they do not qualitatively look at the behavior of the human)

- It is possible to imitate

- Both parties can get better at the task

- It must be easier to be successful at the task when collaborating

- The task must have a physical aspect, allowing interactions to be physically grounded



Different stages of the prototyping proces. The top image was made by Dimitra Chantzopoulou.

All of these aspects are present in the final design of the task as presented in the paper part of this report. Both parties have their own knowledge which is complementary to that of their partner, enabling them to initiate action that their partner cannot initiate. Since the interaction between the robot and the human is very symmetrical, imitation is definitely possible. The knowledge of both partners is relevant for the task, making collaboration beneficial and enabling the partners to learn how to use their knowledge in the best possible way. Last, the communication and coordination between the human and the robot happens in a very physical manner, allowing interactions to be physically grounded but also allowing for very subtle and implicit interactions. This definitely contributed to the richness of the data gathered.

Interaction design

Most of the interaction design for the robot used in the experiment was inspired by the physical pulling that the communication between a guide dog and a blind person consists of. The fact that this interaction can be very subtle but very explicit as well was useful for making sure the leadership shifts could exist on a continuum. The shape of the robot was made to have a front and a back part, so that its movements would have a clear directionality. A black fabric cover was made to give the robot a smooth but non-anthropomorphic look.

In terms of electronics, the robot is made out of a modified remote controlled children's car. On top of that, the insides contain a stretch sensor connected to the leash that can measure how strongly the leash is being pulled. This was initially designed to serve as input for the Wizard controlling the robot, to enable them to adapt their behavior to the leash activity.

Dutch Design Week

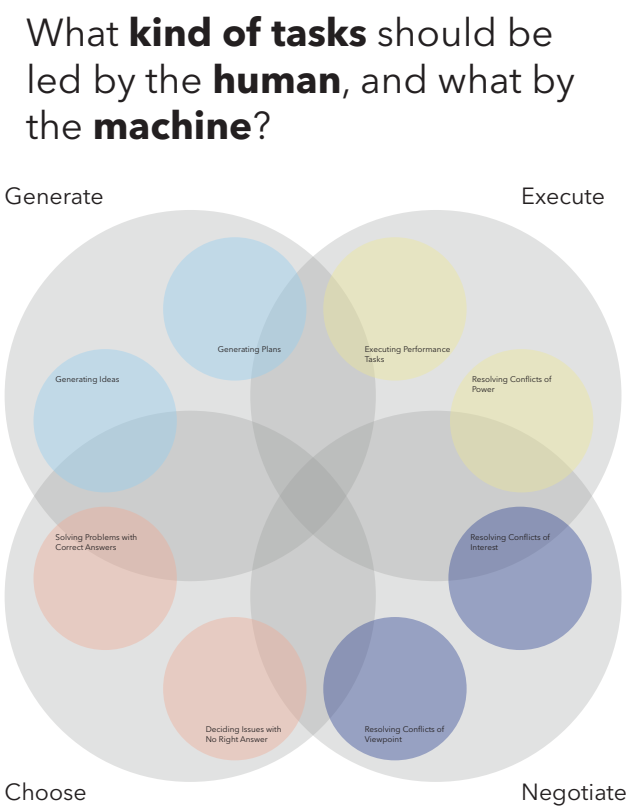
After the M2.1 Demo Day, I was invited to present my project at the Dutch Design Week in October 2019. At first I was hesitant to do this since I did not have a prototype or clear design to show. However, after some conversations with the curator of the exhibition, we decided that the ideas within my project were worth sharing to a more general design audience and fit well within the focus of the exhibition.
The goal of my Dutch Design Week exhibition was to start a conversation with the general public about the idea of human and machine collaborating, as well as to show people why mutual adaptation in human-machine collaborations is an important direction for design and research. In order to achieve this, I first created a small storyline animation that explained with visuals the thinking behind my project. This was presented as a video at the exhibition.
While this video communicated the abstract idea quite well to an interested audience, it would require people passing my stand to watch for quite a while and understand abstract ideas rather quickly. For that reason I decided to create something that attracted attention, made them think and allowed them to actively participate in the discussion. This resulted in three question and growing data visualization posters. I presented three large posters, each with a question, that could be answered by putting a sticker on the poster. The three questions were as follows:

1. What kind of tasks should be led by the human, and what by the machine?

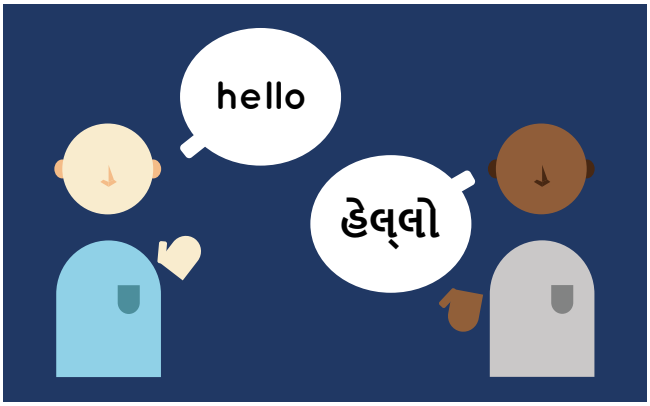
2. What does the machine need to learn and understand about you before collaborating?

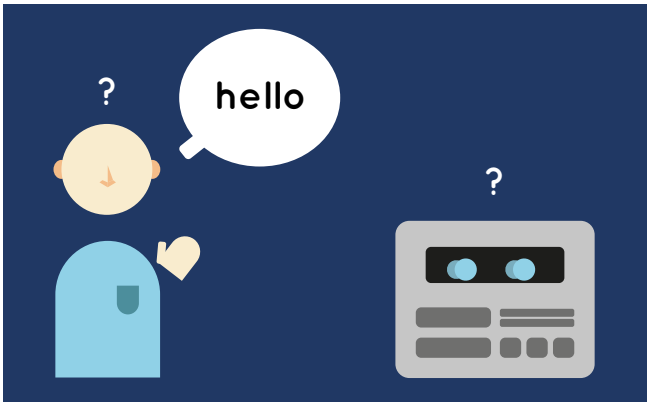
3. What do you need to learn and understand about the machine before collaborating?

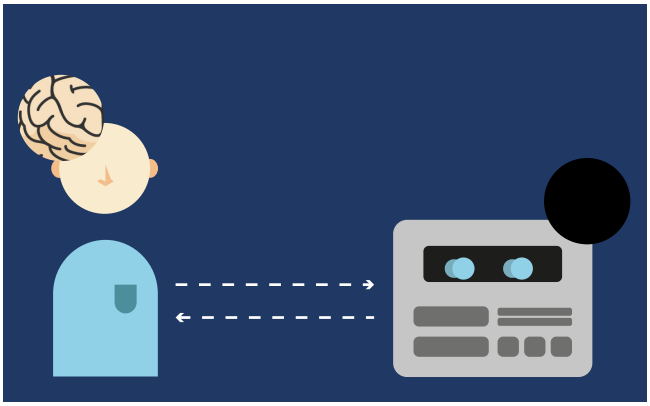
The first question could be answered by putting stickers of little humans and robots in a field with circles based on the task typology by McGrath [6]. The other two questions could be answered by writing a short answer on a blank sticker and pasting it in a field divided up into the four categories Personality/Algorithm, State, Capabilities and Goals. The predefined fields were chosen to give people some guidance in coming up with an answer for the quite abstract and philosophical questions.



The question poster for question 1 and the exhibition stand as presented at the Dutch Design Week.







Several of the illustrations used for the video that was presented during the Dutch Design Week.

Besides the above described materials, I designed a tablet app which showcased three example scenarios in which a human and an intelligent machine need to collaborate. In all of these scenarios, I presented visitors with a dilemma due to conflicting intentions of the human and the machine. They then had to make a choice to follow themselves, follow the machine or make a compromise. The scenarios served to make the idea more concrete and graspable for visitors, as well as to make them think about such dilemmas a bit deeper. My experience during the Dutch Design Week was that overall, the materials I designed communicated the ideas I had very well, as I had many interesting discussions with many different kinds of people. The responses to the sticker posters mainly showed that people have very stereotypical ideas about question number 1, where machines are meant to execute performance tasks and possibly generate plans whereas humans should generate ideas and engage in all tasks requiring social skills. Some people did try to think out of the box, causing some stickers to be placed differently, which in turn caused interesting discussion. For question number 2, people had many demands for what a machine should know about them: basically they feel like the machine should perfectly understand them. How that relates to privacy is an interesting discussion, especially since some stickers expressed the need for a machine to understand when privacy is necessary. The poster with question 3 contained relatively little stickers, which showed me that people do not often realize that they themselves might need to learn about and adapt to the machine as well. From this, I concluded that the research that I am doing and the vision that I have requires a change in mindset. It will be interesting to see how this develops in the future with new technological advancements.

Pilot

Right after the Dutch Design Week, my regular research activities continued and I finalized the setup of the experiment. To test whether the experiment would run as I planned, I conducted a pilot test with two participants. Both participants went through the whole process, to enable me to evaluate every aspect. In general the pilot went better than I expected as the participants enthusiastically did as best as they could. They expressed that they enjoyed the experiment, and already showed some interesting behaviors in terms of leadership shifts. However, there were a few limitations that revealed itself:

- At the moment of pilot testing, there was no explicit feedback for when the participants picked up an object. The robot would simply hold still for a while, and continue its course. While both the behavior as well as the interview answers of the participants showed that they still understood roughly when an object was picked up, they both mentioned that they would have liked to have more explicit feedback. As there was no other explicit feedback about participant's performance, I decided to add a little noise that indicated that an object was picked up in the final experiment.

- In the original set up of the experiment, the prototype measured the pull of the leash with a stretch sensor. This value would be sent to a server so that it could be read out by the Wizard (the controller of the robot), as an aspect that could influence the behavior of the robot. However, the internet in the location of the experiment was not stable enough for the ESP that controlled this process to reliably send out values. During the pilot, however, it became clear that it was quite easy to see when the leash was pulled and when it was loose. Therefore, I decided to not use the sensor.

- Just before the first pilot round, it became clear that the range of the remote controlling the robot was much smaller than expected; therefore, the Wizard had to walk a few meters behind the robot. Because of this, the participant would know that the robot was controlled by a human. Several attempts at increasing the range of the remote control were unfortunately not successful. After reviewing the videos from the pilot test, however, it seemed that the participants did not really let their behavior be influenced by the presence of the Wizard. While of course there might have been some unconscious influence, generally their attention was focused towards the robot and the task, and not towards the Wizard.

- Last, it became clear that the participants assumed that the virtual objects would be placed in the same locations every round. Since this was not the case I changed the instructions slightly to explicitly mention that the locations and number of objects would be randomized every new round.

The camera setup for the experiment with a view on the field and the robot.

Experiment

The final experiment started running after the insights from the pilot were incorporated in the final design. Over a period of three weeks I was able to find 18 participants willing to join my experiment. Overall, everything went as planned. Moreover, in the first week of running the experiment I received official approval for my experiment from the Ethical Review Board of Industrial Design. With the newly setup ethics procedures that were not yet running smoothly, I was happy that I was able to still go through the process of having my experiment design checked for ethics issues.

Stills from the video recordings of the participants during the experiment.

Data Analysis

After running the experiment, I defined a plan for analyzing all the data I had gathered. In the process of defining such a plan, I mostly attempted to create a meaningful balance between finding commonalities between participants and maintaining the richness the video data had to offer.
For the qualitative analysis of the videos, I started with an open coding process to simply describe the behavior that could be observed. This approach however turned out to be limited in giving me an overview of behavior across different rounds. One of the main purposes of my research was to observe how behavior would change over time, hence my experiment design in four rounds. In order to make a judgement about the behavioral development over the rounds, I wanted to view the videos of all four rounds at the same time, which is obviously impossible. Eventually, moving to a closed coding scheme proved to be more successful for this purpose. I created a closed coding scheme derived from the open codes. My closed codes were quite low level and close to the literal behavior of the participants, because I did not want to make interpretations just yet. From these closed codes, the coding program I used (The Observer XT) created visualizations that allowed me to see an overview of different rounds in one screen. This enabled me to indeed describe how the frequency and duration of specific behaviors changed over the different rounds for a specific participant.
After the coding of the videos was done, I was presented with a new challenge. I wanted to describe the overviews of participants that I had in a short but meaningful manner, and link these to the scores of the different questionnaires. I decided to mostly take a clustering approach, where I clustered participants on similar behaviors for different behavioral factors. In this, a cluster could be as small as one participant and as large as all participants, to make sure I did not loose any observed behaviors. Having different clusterings for the same participants also allowed me to compare them between each other to see if there were any overlaps. I did this visually using crosstabulation. Unfortunately not many results came out of the crosstabulation, as the data was quite spread out and the number of participants not very large. I considered to check for significance using a CHI2 test, but decided to not do this because it would most likely not yield any result. The few results that I did get out of the crosstabulation process can be seen as pointers for future research.
Next to the clusters, I also had some quantitative data that could be tested for correlations: the total duration of different behaviors in task execution and the Collaboration Fluency questionnaire scores. After correcting the total duration of behaviors for the actual time it took people to complete the task, it turned out that there was a weak negative correlation between two of the leading behaviors and the Collaboration Fluency score. It would be interesting to see if a single variable for 'leading behavior' could be constructed from the different separate leading behaviors. Currently, I did not have the knowledge or time to do this, but it is something I will keep in mind for future projects.

References Process

[1] T. C. Scott-Phillips, S. Kirby, and G. R. S. Ritchie, “Signalling signalhood and the emergence of communication,” Cognition, vol. 113, no. 2, pp. 226–233, 2009.

[2] K. Youssef, P. R. S. De Silva, and M. Okada, “Sociable Dining Table: Incremental Meaning Acquisition Based on Mutual Adaptation Process,” in Social Robotics, vol. 8755, M. Beetz, B. Johnston, and M.-A. Williams, Eds. Cham: Springer International Publishing, 2014, pp. 206–216.

[3] S. Nikolaidis, D. Hsu, and S. S. Srinivasa, “Human-robot mutual adaptation in collaborative tasks: Models and experiments,” I. J. Robotics Res., vol. 36, pp. 618–634, 2017, doi: 10.1177/0278364917690593.

[4] M. Krüger, C. B. Wiebel, and H. Wersing, “From Tools Towards Cooperative Assistants,” in Proceedings of the 5th International Conference on Human Agent Interaction - HAI ’17, Bielefeld, Germany, 2017, pp. 287–294, doi: 10.1145/3125739.3125753.

[5] I. Konvalinka, P. Vuust, A. Roepstorff, and C. D. Frith, “Follow you, Follow me: Continuous Mutual Prediction and Adaptation in Joint Tapping,” Quarterly Journal of Experimental Psychology, vol. 63, no. 11, pp. 2220–2230, Nov. 2010, doi: 10.1080/17470218.2010.497843.

[6] J. E. McGrath, A typology of tasks. 1984.

General **Conclusion**

Over the past year, I have worked on a topic through which I attempted to bring Industrial Design and Artificial Intelligence together. It was not always easy to balance the two, as the status quo in both research fields differs greatly. The academic field of AI is traditional and theoretical, and human-centered AI work uses controlled experiments that provide conclusive results, but are sometimes far removed from the fuzziness of the real world. The academic field of Design emphasizes this fuzziness, exploring how different real people behave in relation to designs placed very often in real-life environments. I wanted to make use of the strengths of both by allowing for rich interactions that could be qualitatively evaluated, while at the same time using a setup that was controlled enough to draw concrete conclusions and feed into theoretical work.
I believe that conceptually, the work I did met my criteria. The interaction behavior and designed robot and task are concrete enough that it would be possible to start working on an AI model that autonomously portrays the behavior, while possibly even using the insights gained in the experiment to make the behavior of the robot more adaptive and complex. At the same time, people interacting with the robot and task had the freedom to act unexpectedly and break the rules of the experiment without making the results unusable. In terms of execution and evaluation, I believe that there is still room for improvement and learning. The experiment was also a way to let people experience what it would be like to collaborate with a machine. If I had more time, it would have been interesting to elaborate on this experience a bit more, to inform my design and improve on it. Moreover, the combination of the different approaches produced a large amount and variety of data, making it a challenge to condense it into valuable results. Working on processing the results took more time than I planned; in future projects I will schedule more time for a thorough exploration of my data and try to plan the analysis more in advance.
The final semester helped me grow as a researcher in both skills and confidence, to the point that I now have a much better understanding of what my two disciplines can bring me. I do know that there is still a lot of space for me to grow even more, and in the future I will keep exploring how methods and approaches from design can inform research into complex human-AI interactions. I am looking forward to this future!

Acknowledgements

The graduation project was a challenge at times, but several people helped me to make it to the end in such a way that I can be proud of myself.

First of all, I would like to thank Matthias for his critical coaching. You challenged me to think as far as I could, and this definitely increased the quality of my project. Harm, I would like to thank you as well, for assessing me several times but also for your willingness to help me out and give me advice during the last few weeks. Moreover, I would like to thank Emilia, for helping me find a location for my experiment and software for my analysis.

The TU/e Robocupteam also deserves thanks for letting me use their field and for putting up with my stuff being in their spaces during the experiment. It was great to experience another part of the university and see what you guys were working on while I was waiting for a new participant.

Besides all this, this final year would have been much less fun without all the other students sitting in the CDR space. The lunches, the Fika moments and all the funny and useful conversations definitely helped me work more efficiently (and just to relax at other times).

Last, I want to thank Joris for his trust in my capabilities, his cooking projects that provided me with enough fancy food in my freezer, and his love and care. I hope that you are proud of me for bringing another Master thesis to its end!

Appendix A.

Information Sheet and Informed Consent

Subject information for participation in scientific research

Mutual Understanding for Human-Machine Collaboration

Introduction

Dear Sir/Madam,

You are asked to take part in a scientific study.
Participation is voluntary. Participation requires your written consent. Before you decide whether you want to participate in this study, you will be given an explanation about what the study involves. Please read this information carefully and ask the investigator for an explanation if you have any questions. You may also discuss it with your partner, friends or family.

1. General information

In the future, due to artificial intelligence developments, smart computers and robots will get more intelligent and autonomous. Because of this, more and more tasks will arise in which humans and machines will have to collaborate, making use of the strengths of both. This research investigates the interaction between humans and machines in such tasks.

2. Purpose of the study

This study is an exploratory investigation of a situation in which human and machine have complementary capabilities. The purpose of the study is to observe how human and machine adapt their behaviour to accommodate this, as well as to understand how this process influences understanding of and trust in the machine.

3. What participation involves

You will need to navigate to a specified target in the room together with the machine. In the environment, there are several virtual objects of varying value of which the machine knows the location. The machine does not know the target location. You can gather the virtual objects, which give extra points, but you also need to navigate to the target location as soon as possible because the time to reach it also counts towards your final score. Together with the machine you need to find out the best way to complete the task with as many points as possible.
You will perform the task four times, each time with randomly allocated virtual objects.
During the study, data is collected about your behaviour by video recording the process. After each time you perform the task, we will let you complete a questionnaire about your understanding of and trust in the machine.

4. What is expected of you

In order to carry out the study properly it is important that you do your best to finish the task at the highest score possible at all times. The participant with the highest score on a single run will win a gift card of €10,-.

It is important that you contact the investigator:

- if you no longer want to participate in the study.
- if your contact details change.

5. If you do not want to participate or you want to stop participating in the study

It is up to you to decide whether or not to participate in the study. Participation is voluntary. If you do participate in the study, you can always change your mind and decide to stop, at any time during the study. You do not have to say why you are stopping, but you do need to tell the investigator immediately.
The data collected until that time will still be used for the study.

If there is any new information about the study that is important for you, the investigator will let you know. You will then be asked whether you still want to continue your participation.

6. End of the study

Your participation in the study stops when

- you choose to stop
- the experiment has been successfully executed
- the investigator considers it best for you to stop
- the government or Ethical Review Board, decides to stop the study.

The study is concluded once all the participants have completed the study.

7. Usage and storage of your data

Your personal data will be collected, used and stored for this study. This concerns data such as your name and date of birth. The collection, use and storage of your data is required to answer the questions asked in this study and to publish the results. We ask your permission for the use of your data.

Confidentiality of your data To protect your privacy, your data will be given a code. Your name and other information that can directly identify you, will be omitted. Data can only be traced back to you with the encryption key. The encryption key remains safely stored in the local research institute. The data that is shared for assessment and publication will only contain the code, not your name or other data with which you can be identified. The data cannot be traced back to you in reports and publications about the study.

Access to your data for verification

Some people can access all your data at the research location. Including the data without a code. This is necessary to check whether the study is being conducted in a good and reliable manner. Persons who have access to your data for review are the researcher, the supervisor of the researcher, and one assistant appointed by the researcher. They will keep your data confidential. We ask you to consent to this access.

Retention period of your data

Your data must be kept for 5 years at the research location.

Withdrawing consent

You can withdraw your consent to the use of your personal data at any time. The study data collected until the moment you withdraw your consent will still be used in the study.

More information about your rights when processing data

For general information about your rights when processing your personal data, you can consult the website of the Dutch Data Protection Authority.

If you have questions about your rights, please contact the researcher.
If you have questions or complaints about the processing of your personal data, we advise you to first contact the research location. You can also contact the Data Protection Officer of Eindhoven University of Technology or the Dutch Data Protection Authority.

8. Any questions?

If you have any questions, please contact Emma van Zoelen (e.m.v.zoelen@student.tue.nl).
If you have any complaints about the study, you can discuss this with the investigator. If you prefer not to do this, you may contact the supervisor Matthias Rauterberg (g.w.m.rauterberg@tue.nl).

9. Signing the consent form

When you have had sufficient time for reflection, you will be asked to decide on participation in this study. If you give permission, we will ask you to confirm this in writing on the appended consent form. By your written permission you indicate that you have understood the information and consent to participation in the study. The signature sheet is kept by the investigator. Both the Investigator and yourself receive a signed version of this consent form.

Thank you for your attention.

Subject Consent Form

Mutual Understanding for Human-Machine Collaboration

I hereby declare that:

- I have read the subject information form. I was also able to ask questions. My questions have been answered to my satisfaction. I had enough time to decide whether to participate.
- I know that participation is voluntary. I know that I may decide at any time not to participate after all or to withdraw from the study. I do not need to give a reason for this.
- I give permission for the collection and use of my data to answer the research question in this study.
- I know that some people may have access to all my data to verify the study. These people are listed in this information sheet. I consent to the inspection by them.

- I want to participate in this study.

Name of study subject:
Signature: Date: __ / __ / __

I hereby declare that I have fully informed this study subject about this study.

If information comes to light during the course of the study that could affect the study subject's consent, I will inform him/her of this in a timely fashion.

Name of investigator (or his/her representative):
Signature: Date:__ / __ / __

Appendix B.

Collaboration Fluency and Personality Questionnaires

Participant Number

Date

Time

The human-robot team worked fluently together.

The human-robot team's fluency improved over time.

The robot contributed to the fluency of the interaction.

I had to carry the weight to make the human-robot team better.

The robot contributed equally to the team performance.

I was the most important team member on the team.

The robot was the most important team member on the team.

I trusted the robot to do the right thing at the right time.

The robot was trustworthy.

The robot was intelligent.

The robot was committed to the task.

The human-robot team improved over time.

The robot's performance improved over time.

I feel uncomfortable with the robot.

The robot and I understand each other.

I believe the robot likes me.

The robot and I respect each other.

I am confident in the robot's ability to help me.

I feel that the robot appreciates me.

The robot and I trust each other.

The robot perceives accurately what my goals are.

The robot does not understand what I am trying to accomplish.

The robot and I are working towards mutually agreed upon goals.

I find what I am doing with the robot confusing.

The robot has had an important contribution to the success of the team.

The robot was committed to the success of the team.

I was committed to the success of the team.

The robot was cooperative.

Based on Hoffman, G. (2019). Evaluating fluency in human–robot collaboration. *IEEE Transactions on Human-Machine Systems*, *49*(3), 209-218.

Participant Number

Date of Birth

Current Date

I am the life of the party

I sympathize with others' feelings

I get chores done right away

I have frequent mood swings

I have a vivid imagination

I don't talk a lot

I am not interested in other people's problems

I often forget to put things back in their proper place

I am relaxed most of the time

I am not interested in abstract ideas

I talk to a lot of different people at parties

I feel others' emotions

I like order

I get upset easily

I have difficulty understanding abstract ideas

I keep in the background

I am not really interested in others

I make a mess of things

I seldom feel blue

I do not have a good imagination

Based on Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192.