university of groningen / faculty of mathematics and natural sciences / artificial intelligence

Backchanneling in Human-Robot Interaction

Differences in Human Backchanneling Behavior when Communicating with another Human versus a Robot

Adna Bliek

Internal Supervisor: Prof. Dr. N.A. Taatgen
(Artificial Intelligence, University of Groningen)
External Supervisor: Prof. Dr. T. Hellström
(Umeå University, Sweden)

Artificial Intelligence
University of Groningen, The Netherlands

Abstract

During human conversations, information is not only conveyed by the speaker but also by the listener. For this, the listener is using backchannel feedback. Backchanneling can be done verbal or non-verbal and is a key aspect of human conversation, it shows listener attention and contributes to a more fluent interaction. In recent years, the effect of a robot's backchanneling while listening to a human has been investigated. But the effect of backchanneling by a human when listening to a robot has been studied poorly.

In this thesis, we investigate (1) how a robot's behavior affects the human's backchanneling feedback, (2) whether the amount of backchanneling feedback has an effect on how the conversation is perceived, and (3) how human backchanneling behavior differs when listening to a robot compared to a human. The investigation of human backchanneling is important in Human-Robot Interaction to create fluent conversations and to be able to use the feedback of the human listener to adjust the conversation.

We conducted an experiment looking at the backchanneling cues gaze, gesture, and pauses that were exhibited by a semi-humanoid robot. We found that pauses have a significant influence on the backchannel behavior of the human. Furthermore, we found that backchanneling behavior when listening to a robot or a human differs significantly.

Acknowledgments

I would like to thank everyone that helped me in the process of writing this theses. First of all I would like to thank my supervisor Niels Taatgen who always gave me input when needed. Second, I would like to thank Thomas Hellström and Suna Bensch who supervised me in Umeå and gave me the great opportunity to come to Sweden and to work with them and the Pepper robot. They were a great help designing the experiment and discussing the results with me to improve the pilot experiments. I would also like to thank them for the opportunity of writing a paper for the RO-MAN conference together. I would also like to thank my family and friends who made me the person that I am today and helped me to successfully finish University and to write this thesis. Finally, I would like to thank my friends in Umeå who made Umeå feel like home, participated in my experiments and helped me improve the experiments with their feedback.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Robots are nowadays not only used in traditional industrial contexts but more and more also in interactive applications with humans. They are being used as personal assistants, caregivers, companions and teachers. During the corona epidemic they have been used as caregivers and entertainment in hospitals, and as reminders to keep the social distance in supermarkets and other public spaces [1].

With growing complexity of the tasks that robots can carry out, the expectations of the communication skills of the robot grow. Humans expect to have human-like conversations with robots, even if the robot was not designed for that [17]. One part of human-like conversation is backchanneling. In linguistics, backchannelling refers to social cues given by the listener that encourage the speaker to go on [15]. Backchannelling can be done using verbal and non-verbal cues. Examples of verbal backchanneling are phrases like "yeah", "yes" and "mmh". Examples of non-verbal backchanneling are gestures, facial expressions, and nods.

1.1 Human Communication

Communication between humans is a complex, multi-modal process. We do not only use verbal messages to convey the meaning of our utterances, but also non-verbal messages. Both the verbal and non-verbal messages have to be received and understood by the listener to fully understand the message the speaker is trying to convey [27]. Non-verbal messages can be present in a wide range of modalities, examples are facial expressions, gestures, nods, and body posture.

Another aspect of communication that makes it more complex, is the cooperative effort of the speaker and listener. Communication is not a one-way street where the speaker is speaking and the listener only listening. Grice stated that "Our talk exchanges do not normally consist of a succession of disconnected remarks, and would not be rational if they did. They are characteristically, to some degree at least, cooperative efforts" [11, p.26]. This statement can explain a need for feedback that is given by the listener to make the conversation a cooperative effort, as well as to make a continuous conversation out of the statements given by the speaker.

[1] Z. Thomas, "Coronavirus: Will Covid-19 speed up the use of robots to replacehuman workers?", BBC(Apr. 2020), `https://www.bbc.com/news/technology-52340651` (accessed: 26-06-2020)

1.2 Significance of the Study

Backchanneling has been studied in human-human conversations in different cultures, situations, and modalities [12, 24, 8, 21]. In the recent years there has also been research in the field of backchanneling in Human-Robot Interaction(HRI). This research has been mainly focused on studying backchanneling in HRI in the direction of a robot giving backchanneling feedback while listening to a human [13, 10, 16]. The focus was on how a robot could communicate that it was listening to the human and create a more fluent conversation. In this research, we look at how to trigger backchanneling behavior by a human when listening to a robot. This part of HRI is largely missing in the literature. With this research we hope to contribute preliminary results into how human backchanneling affects HRI and how it can be triggered by a robot, so that a robot can actively contribute to more fluent and effective conversations.

1.3 Research Question

The main research questions addressed in this thesis are:

1. *How should a robot behave in order to trigger a human's natural backchanneling behavior?*

2. *Does human backchanneling affect how the listener perceives the interaction?*

3. *How does a human's backchanneling behavior differ when listening to a robot compared to a human?*

In order to answer the questions, an experiment with human participants listening to a story-telling robot was conducted. The first research question was investigated by implementing three backchanneling-inviting cues in a robot. The cues were *pauses*, *gestures* and *gaze*. The effect on the human backchanneling behavior was measured by counting the amount of backchanneling feedback given by a human during each condition. The second research question was examined by looking at how the participants self-report their perception of the conversation. The third research question was looked into by comparing the backchannelling behavior of the participants while listening to a story told by the robot versus a human.

1.4 Overview of the Thesis

Three experiments were conducted, all using the semi-humanoid robot Pepper. The first two experiments were pilot experiments. The first experiment was inspired by a map task [1]. During the second and third experiment, ethical dilemmas were told by the robot. To answer the first research question, three backchanneling cues were used by the robot: *gestures*, *gaze* and *pauses*. We found that pauses have a significant influence on the amount of backchanneling performed by the human participants when listening to the ethical dilemmas. The other two backchanneling cues did not have a significant influence on the amount of backchanneling. To answer the second research question, we looked at ratings by the participants, and did not find a significant effect of backchanneling on the perception of the conversation. The third research question was answered by comparing

the amount of backchanneling by the participants to the robot versus a human. We found that the participants backchanneled significantly more to the human and used different backchanneling behaviors.

1.5 Outline

The remainder of the thesis is structured as followed: First, in Chapter 2 the background is discussed, describing backchanneling research in human-human communication, and the current state of the art in relevant parts of human-robot interaction. Then, Chapter 3 discusses methodology, describing which robot and methods have been used in the experiments. Furthermore, it describes the two pilot experiments and the design and implementation of the final experiment. Chapter 4 describes the results obtained in the final experiment. Chapter 5 discusses the results and puts them into the context of the background. Chapter 6 concludes the thesis by describing possible future research and summarizing the research.

Chapter 2

Background

2.1 Backchanneling Definition

Verbal conversations can be defined by at least two persons that convey information to each other using verbal utterances. In addition to the verbal content of the conversation, non-verbal information is present in face-to-face conversations.

During most conversations, one participant takes the role of the speaker for a time while the other participant is listening. Even though the role of the listener may seem more passive, the listener still uses verbal, para-verbal and non-verbal messages to convey information to the speaker. This feedback by the listener is called backchanneling. The concept of backchanneling is a well-known linguistic concept that was first described by Yngve in 1970. In [31], they describe how the speaker communicates on the main-channel, and the listener gives minimal non-interrupting feedback on the back-channel. Examples of verbal backchanneling include phrases like 'uh-huh', 'yes', 'yeah', and 'mmh'. Examples of non-verbal backchanneling include nodding, shaking the head, and facial expressions.

In this thesis, we will define backchanneling as in [29] using three clauses to characterize backchanneling.

1. The feedback responds directly to the content of an utterance of the speaker
2. The feedback is optional
3. The feedback does not require acknowledgment by the other

The first clause makes sure that feedback is related to what the speaker is saying and not to stimuli outside the conversation. The second and third clause ensures that the feedback given is not a response to a question or a change of speaking turn. The backchanneling feedback does not require acknowledgment by the speaker, even though feedback can be given. Besides, backchannel feedback should not interrupt the speaker and the listener should not take over the speaking turn [4].

2.2 Backchanneling in Human-Human Communication

Backchannel responses are a universal phenomenon and can already be observed with six-year-old children, but children use backchannel responses less frequently than adults [7]. Backchannel responses can be used for a variety of reasons by the listener. Maynard [25]

has defined six categories of reasons why listeners use backchannel feedback. The defined backchannel categories are continuer, display of understanding, support towards the speaker's judgment, agreement, strong emotional response, and minor addition, correction, or request for information. The most frequent function found in English conversations is as continuer [25]. Continuers are used by the listener to communicate that they do not intend to take over the speaking turn but would like the speaker to continue.

The categorization of backchannel behaviors can be challenging as words can be ambiguous in meaning especially when only looking at a single word. For example, the often-used backchannel 'yeah' can be categorized as a continuer, display of understanding, or support towards the speaker's judgment in a backchannel situation or even as a response to a question in a non-backchannel situation. To distinguish between the different meanings lexical and semantic knowledge has to be used [18].

2.2.1 Types of Backchanneling

What kind of backchanneling response is used varies between languages, cultures, and individuals. In addition, the used backchannel responses are also dependent on the context, whereas *nodding* is often used in face-to-face conversations, it does not yield any feedback in phone conversations.

Backchanneling feedback can be divided into two categories: generic and specific feedback. Generic backchannel feedback includes all feedback that is not specific to the context and is often seen as the *standard* backchannel response. Generic responses include nodding or verbal phrases like 'yeah', these responses can most often be characterized as *continuer*. The second category is specific backchannel responses. These responses are related to the content, for example, looking sad at appropriate moments, or mirroring the speaker's gestures and movements. These responses permit the listener to become the co-narrator, illustrating, or adding to the story [3].

2.2.2 Backchannel Timing

Backchannel feedback can be given during pauses or overlapping the speech of the speaker. Feedback during speech pauses is a universal phenomenon, whereas the presence of backchannel feedback overlapping speech is not present in all cultures and languages. Native German speakers produce less overlapping backchannel responses than American speakers in their native languages and their conversation can be disturbed by overlapping feedback [12].

In American English, it was found that backchannel responses seem to follow intonational phrases with raising pitch [5]. Besides, it has been found that in American English backchanneling is being done at grammatically significant breaks, such as at ends of clauses or sentence-final positions [25].

The amount of backchanneling responses of distracted listeners was found to be significantly lower than that of listeners attending to the content of the story. This effect was found for both generic and specific responses but was larger for specific responses. An increase in the cognitive demand of a task while still being able to attend to the story did not change the amount and kind of backchanneling feedback of the listener [3].

2.2.3 Effect of Backchanneling Feedback

The quality of how well the speaker can tell a story can be influenced by the amount of appropriate backchanneling. When a listener is distracted and does not use as many backchannel responses and less specific responses, it was found that the quality of the storytelling was lower than when appropriate backchanneling was present [3]. Besides, the number of backchanneling responses was also found to have a weak positive correlation with task success [5]. But more backchanneling does not always have to be better, too much backchanneling has for example been found to have a negative effect on the enjoyment of the conversation [24].

In addition to improving the storytelling, backchanneling can also express rapport and grounding in a conversation [27]. Rapport is the perceived connection between the listener and speaker and grounding can show acknowledgment of the listener.

2.2.4 Effect of Backchanneling Cues

As described before in subsection 2.2.2, the amount of backchanneling can be influenced by different factors, for example backchannel cues. Backchannel cues are cues of the speaker while speaking provoking a backchannel response by the listener. Such a cue can, for example, be a rising pitch, a pause, a gaze shift, or a gesture. These cues can have an individual influence on the backchannel behavior, but they can also have a joint influence. a joint influence is the influence of multiple cues used at the same time is bigger than the individual influences of the used cues [27]. The cues also do not always have to be synchronized with the speech, a gesture can, for example, be performed a bit after the corresponding speech[26].

In this thesis, the effect of the backchanneling cues pause, gesture, and gaze will be investigated.

Pause

In a conversation different kinds of pauses can be present. Pauses can be categorized by function and whether they are filled or silent. Filled pauses are, in contrast to silent pauses, pauses that are filled with filler words like 'eehm'. The specific filler words can vary per culture and language. Three functions of pauses can be identified: a psycholinguistic function to allow the speaker to breathe, a cognitive function to allow the speaker to plan the next part of their speech, and a communicative function, to help the listener to identify significant syntactic places is the speech stream [6]. The last two functions can indicate to the listener whether the speakers want to continue their turn or give the turn to the listener.

Gesture

Gestures are a universal phenomenon, there has not been any report of a culture that is not using gestures accompanying their speech [20]. They are already present and synchronized in children in the one-word stage. Those children will, for example, gesture to an object while uttering the word *drink* to show what they would like to drink [9]. Even though gestures are present in all cultures, they vary between them in terms of position, size, and plane (lateral, sagittal, or vertical) [20]. A gesture that is appropriate in one culture can be misunderstood in another or even be perceived as rude.

Gaze

For how long and how often someone looks at the other during a conversation varies greatly between individuals [27]. Additionally, it also varies according to the conversation partner, as it is closely related to the behavior of the other. A difference can also be found between the listener and the speaker, people look less at the other while talking than while listening. The listening part looks for long times at the other with short breaks in between, whereas the speaking part alternates between looking at the other and away from them. Mutual gazes tend to be short. The gaze behavior of the listener can give non-verbal information about them willing to take the turn and the speaker usually has a characteristic head posture and gazes at the listener before stopping their turn [19].

2.2.5 Cultural Differences

The use of verbal and non-verbal backchanneling seems to be a universal feature of human communication. But specific backchannel behaviors, frequency, distribution, and function of backchanneling behavior are particular to certain languages and/or cultures [12].

It was, for example, found that native Japanese listeners use more synchronized head movements and overall more backchanneling feedback than American listeners. Furthermore, Japanese listeners tend to use more overlapping backchanneling than native American English speakers [25]. In another study, comparing native German and native American English speakers, it was found that the German speakers produce less backchanneling responses and less overlapping responses. Moreover, the two groups differed in terms of what type of conversational function the backchanneling responses had [12].

When learning a second language also new appropriate backchannel behavior has to be learned. Proficient second language speakers switch between appropriate backchannel behavior when changing language. For example, native Chinese Mandarin speakers do not use the in Mandarin often-used phrase *'is that so'* (*Shi Ma*) in English [24]. On the other hand, native German speakers who are bilingual in American English were found to be using more backchannel responses in German than non-bilingual native German speakers and more overlapping backchannel responses [12].

Swedish

Swedish shares linguistic characteristics with other Scandinavian languages, which are often researched together.

Fant [8] researched how the backchanneling behavior in Scandinavian and Hispanic cultures differ. They suggest that Scandinavian conversation patterns are more similar to English patterns than to Hispanic. They found that Swedes do not frequently interrupt each other, and thus backchannel more during speech pauses of the other. Backchanneling is done mainly using verbal cues that show attention and understanding. Gaze is not a frequently used backchannel cue. Furthermore, Swedes have restrictive norms for turn-taking, which are manifested in the fact that Scandinavian conversational culture can be characterized more as "floor-giving" than as "floor-taking". This means that the speaking turn most often goes to the listener from the speaker, and the listener does not take their turn by interrupting the speaker. For turning over the turn to the next speaker, gaze cues are often used.

2.3 Backchanneling in Human-Robot Interaction

The effects of human backchanneling behavior in Human-Robot Interaction(HRI) have not been extensively studied. In HRI it has mainly been studied how the backchanneling behavior of a robot affects the human but not the effect of backchanneling-inviting cues provided by the robot.

HRI experiments can be done with physical robots or virtual agents, but the results found in virtual agents cannot always be translated to interactions with physical robots.

2.3.1 Simulated Agents

Agent Giving Backchannel Cues

In an experiment by Hjalmarsson and Oertel [13], a virtual agent used gaze as a backchannel cue. The participants were divided into two groups: the agent looked at the participants of the first group at backchannel inviting moment, and at random moments for the second group. The researchers found that the participants backchanneled more when the agent used the backchanneling inviting gaze condition. A limitation of the study that should be noted is that it was not looked at natural backchannel behavior but the participants were asked to press a button when they thought that a backchannel would be appropriate. This experiment shows that the backchanneling inviting cue *gaze* of a simulated agent could influence the behavior of the human listener but more investigation into whether this is also translatable to natural backchannel behavior, and physical robots is needed.

Agent Giving Backchanneling Feedback

Gratch et al. [10] created a virtual agent to create rapport in a conversation of a human with the agent. To create good rapport, the agent used backchanneling behavior when the participant was talking. The feedback was generated using real-time analysis of the acoustic properties of the speech and the speaker's gestures. To test the agent four conditions were tested: human-human face to face, mediated (virtual agents movements were copied from a human), responsive (agent reacts to the participant using the automatically generated responses), and non-contingent (responsive feedback from the earlier session was used, which was not synchronized to the speech of current participant). They found that the responsive agent was as effective as a human listener in creating rapport and more effective than the mediated or non-contingent agent.

2.3.2 Physical Robots

Robot Giving Backchannel Feedback

Inden et al. [16] created five strategies to create models for the best timing of backchannel behavior when listening to a human speaker. The strategies tested were: (1) copying the timing of the original human listener, (2) producing backchannels at randomly selected times, (3) producing backchannels according to high-level timing distributions relative to the interlocutor's utterance and pauses, (4) according to local entrainment to the interlocutors' vowels, or (5) according to both. They concluded that the strategies to generate backchanneling behavior using empirically derived global timing distributions

were perceived as missing fewer opportunities for backchannel feedback than the random strategy.

Hussain et al. [14] used a Markov decision process to train a social robot to backchannel at appropriate times to maximize the engagement of the user. They found that reinforcement learning could be a useful way of learning backchannel behavior as the robot is able to learn immediately from the earlier time-steps and adjust its feedback. But it should be noted that their results have not been tested in an HRI experiment.

Robot Detecting Backchanneling Behavior

Backchannel feedback can be a predictor of how engaged the listener is during a conversation. Lala et al. [22] used this property of backchannel feedback to detect how engaged the human was during a conversation and used the feedback to keep the user engaged during the conversation with the robot.

Chapter 3

Method

In order to answer the defined research questions, three Human-Robot Interaction experiments were conducted using a Pepper robot. The general idea was to have the Pepper robot talk to the human participant, while performing backchannel-inviting cues: gestures, pauses, and gaze movements. Two pilot experiments were first conducted, with the purpose of finding a task that facilitates backchanneling feedback, and to find good settings for the backchanneling-inviting cues. The first pilot experiment used a map task where the participants had to remember a path that was explained by the robot. The task was changed after the first experiment as it did not yield the expected results. For the second pilot experiment and the third experiment the robot told ethical dilemmas to the human participants. The interactions were video recorded such that the backchannel behavior could be analyzed in detail afterwards.

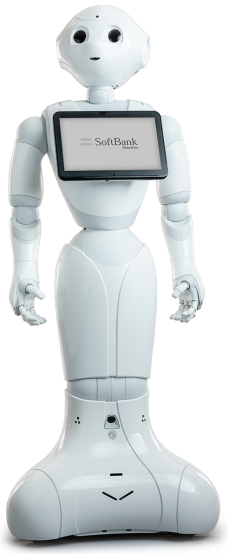


Figure 3.1: The Pepper robot has been used in this experiment. It has a mounted tablet on its chest [1].

3.1 Robot - Pepper

Pepper is a semi-humanoid robot that has been developed by Softbank Robotics [1]. The robot has been created to interact with humans and is currently used in research, as a greeter in businesses and conferences, and in healthcare setting for entertainment of the patients and as support of the healthcare staff.

Pepper stands 1.20m tall and can use a variety of different sensors to interact with and recognize humans. To interact with humans around it, Pepper can use speech, gestures, LEDs and its tablet. To recognize people and its surrounding, Pepper has two 2D cameras and one 3D sensor on the front of its head and four microphones on the top of the head. Pepper has two infra-red sensors and a laser located in the front of the lower part of the robot for navigation and obstacle avoidance. Furthermore, the robot has tactile sensors on its head and hands, and three bumpers close to the ground.

The robot can be programmed in different programming languages using the NAOqi framework which has been developed by Sofbank Robotics. In addition to using external

[1] `https://www.softbankrobotics.com/`

Backchanneling Condition	Pause	Gaze	Gesture
C1	0	0	0
C2	0	0	1
C3	0	1	0
C4	0	1	1
C5	1	0	0
C6	1	0	1
C7	1	1	0
C8	1	1	1

Table 3.1: During the experiments eight different combinations of backchanneling cues were executed by the robot. For example, during condition 6 pauses and gestures are present but no gaze.

Category	Labels
Lean	*neutral*, towards, away
Brow	*neutral*, raise
Smile	*none*, smile
Frown	*none*, frown
Nod	*none*, nod
Shake Head	*none*, shake head
Head Movement	*neutral*, forward, up, tilt
Utter	*none*, utterance
Hand in Face	*not in face*, in face

Table 3.2: List of annotated human backchannel behaviors. The labels show the possible states for each behavior. The first label is the default value, i.e. the value that is expected to be seen when the participant is not backchanneling.

programming languages, Pepper can also be programmed and controlled using the Choreograph Suite which is a block based programming environment. The tablet located on Pepper's chest is an android tablet, android application can be installed or it can be controlled using the NAOqi framework. In the later case an HTML website will be displayed and the feedback from this website can be used by the robot.

In this project, Pepper was programmed using the NAOqi framework for python. In addition, HTML, JavaScript and CSS were used to program the website shown on the tablet. In the JavaScript code the NAOqi SDK was used to establish communication between the python and JavaScript code.

3.2 Backchannel Cues

During all experiments three backchannel cues that can be created by Pepper were manipulated: pauses, gestures and the gaze. The cues were tested in all possible combinations (see Table 3.1). The backchannel cues were placed at points where a topic in the text was finished (see Fig. 3.4c and Fig. 3.5).

Speech pauses were created by adding additional pauses to the text spoken by Pepper. During all three experiments pauses were presented at all possible cue places. After the pilot experiments, the length of the pauses was reduced from 2 seconds to 1.2 seconds as the pauses were perceived as too long by the participants. The gestures that were used were predefined gestures in the NAOqi framework. They were categorized under the topic "explain" and were randomly chosen from the topic. Gestures were performed while the robot was talking. During the execution of a gesture, Pepper moved its body, arms and hands. The gestures were randomly presented 50% of time during the pilot experiments to ensure that the use of gestures felt natural. Using gestures the whole time could have felt unnatural. The amount of gestures was increased to 60% after the pilot experiments. When the gaze cue was used, Pepper did not look at the human participant all the time but would look away and back at the participant at the time when also pauses could be placed. During the pilot experiments, the robot looked away or back at the participant during possible cue places 50% of the time. After the pilot experiments this number

Experiment	Pause Length	Pause Probability	Gesture Probability	Gaze Angle	Gaze Probability	Voice Speed
Map Task Pilot Study	2 sec	100%	50%	yaw: 0.2 pitch: 0.1	50%	slow: 70% normal: 90%
Dilemmas Pilot Study	2 sec	100%	50%	yaw: 0.2 pitch: 0.1	50%	90%
Dilemmas Experiment	1.2 sec	100%	60%	yaw: 0.1 pitch: 0.1	60%	80%

Table 3.3: Parameters of the cue conditions exhibited by the per experiment. The parameters of the cue conditions were changed after each of the pilot experiments according to feedback given by the participants to create a more natural feeling conversations with the robot.

(a) Lean Towards (b) Lean Back (c) Smile (d) Brow Raise

(e) Head Neutral (f) Head Up (g) Head Down (h) Frown

Figure 3.2: Example Images of Backchanneling Behavior

was increased to 60% The parameters of the cue conditions were changed after each of the pilot experiments according to feedback given by the participants and observations made during the experiments. The parameters for each cue condition during the three experiments can be seen in Table 3.3.

3.3 Backchannel Behaviors

The participants' backchannel behaviors that were annotated can be found in Table 3.2. The choice of behaviors was based on the earlier work in [23], and was extended with commonly observed backchannel behaviors in the videos of the pilot studies. The added behaviors were *Shake head* and *Head move*. Examples of real backchannel behaviors can be found in Figure 3.2.

Figure 3.3: Example of a question asked to the participant after each map or dilemma. The question was shown on the tablet of the robot and answered using the touch screen.

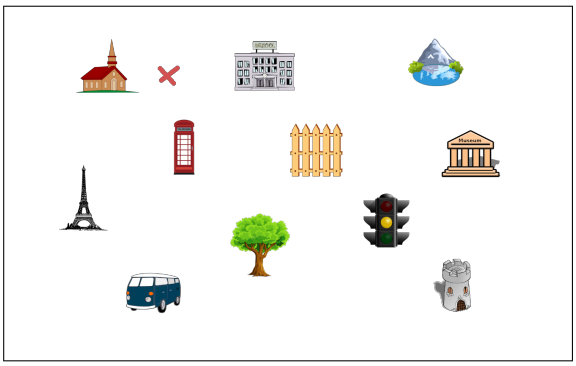
3.4 First Pilot Study

During this pilot study a similar experiment as the one conducted by Anderson et al. [1] was used. The participants were placed in front of the robot and video-recorded using a camera situated slightly behind the robot. The experiment started with the robot explaining the experiment to the participant using a random backchanneling cue combination. The parameters used during the different cue combinations can be seen in Table 3.3. After explaining the experiment, the robot explained an example path to the participant. The participant was shown the explained path on the tablet during the explanation of the example path. This explanation was added to familiarize the participant with how the robot explained a path. After explaining the example path, the experiment started. The robot explained eight different paths to the human participant during the experiment. During the explanation of the path, the participant would see the map with a red cross indicating the starting point on the tablet of the robot, see Figure 3.4a for an example map. The participants were asked to remember the path and draw it on the tablet after the robot had explained the whole path. After each path the participant was asked to answer four questions on the tablet about the conversation with Pepper. The participants ranked the overall conversation, enjoyment, understandability and engagement on a 5-point Likert scale (see Figure 3.3 and Appendix B).

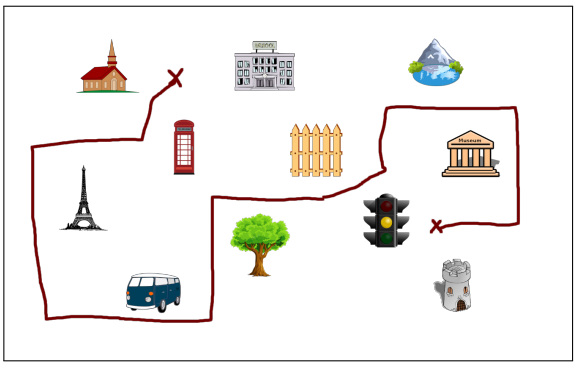
3.4.1 Findings and Conclusions

Participants were recruited at the department of Computing Science at the University of Umeå in Sweden. In total three participants participated in the pilot study after which the pilot study was stopped due to the fact that no backchanneling could be seen during the experiments. After completing the experiment, the participants were asked how demanding they found the task. They all reported that they had to concentrate on the task and the map to remember the task. As a consequence, they looked less at the robot and used the pauses during the speech of the robot to retrace the path in their mind while looking at the tabled instead of looking at the robot's face or movements. This fixation on the screen, instead of having a more relaxed posture and having the possibility to look at the robot, can probably explain the lack of backchanneling found during the experiment.

In addition to finding the task taxing, they also found the way of explaining paths unnatural due to the fact that there were no roads on the maps and they were not placed in the actual world. The last fact made it harder for them to imagine the path.



(a) Map shown to the participants during the experiment. The red cross symbolizes the starting position.



(b) Path that has been explained by the robot. The starting and end positions are marked with a red cross.

Robot: You start on the right of the church; walk down until you are at the left of the telephone box; turn to the west; walk until you pass the tower; walk straight down and turn towards the bus; pass under the bus; turn up after the bus; and walk until you pass the tree; turn towards the stoplight; pass between the fence and the stoplight; turn upwards until you are under the lake; turn to the east and pass between the museum and the lake; after the museum turn downwards; when you passed the museum turn towards the stoplight; walk until you are next to the stoplight; you reached the goal point

(c) Text used by the robot to explain the path. Each semicolon represents a possible moment for a backchanneling cue.

Figure 3.4: Map that has been used during the experiments showing one example starting point and path.

3.5 Second Pilot Study

We concluded from the first pilot study that a task which needs less mental effort and no multi-tasking(like looking at a screen, remembering, mentally drawing a path) would be better suited. Having Pepper explaining ethical dilemmas seems like a less mentally taxing task since the participants then do not have to retain as much information, and since the task of telling stories feels like a more natural task for a robot.

Participants were placed in front of the robot as during the first pilot study, see section 3.4. The robot then verbally explained the experiment to the participant using a random backchannel cue combination. The robot verbally presented eight stories, in random order, to the participant during the experiment. The topics of the stories were ethical dilemmas such as *the trolley problem*. During each story a random cue combination C1-C8 was used, to see whether this triggered human backchanneling. After each story, the robot asked a question related to the story. The participant replied verbally to the question. The end of the reply was determined by the test leader who was located in the same room as the participant and robot during the pilot experiments. Afterwards, the participant rated the conversations using the same four questions as during the first pilot study. In addition to the eight stories told by the robot, one story was presented to the participant by the test leader. This story was randomly told before or after the eight stories presented by the robot. After the experiment the participant was interviewed by the test leader to gather information about how the participants experienced the

Robot: Okay, so you stand next to some railway tracks. And you see a runaway trolley barreling down the tracks.; Ahead, on the track are five people tied up and unable to move.; The trolley is heading straight for them.; You are standing next to a lever and if you pull it, the trolley will switch to a different set of tracks.; However, there is one person on the sidetrack. If you pull the lever, the person on the sidetracks will die.; If you do nothing the five people on the tracks will die.
So what do you do?

Figure 3.5: Example dilemma used during the dilemma pilot study. Each semicolon represents a possible moment for a backchanneling cue.

experiment.

3.5.1 Findings and Conclusion

Participants were again recruited at the Computer Science department at the Umeå University. In total eight participants were recruited, one of them was removed before analyzing the data as they had problems following the dilemmas. Three of the remaining participants were female and four were male.

The video recordings of the experiments were analyzed for three different backchanneling behaviors: nonverbal backchanneling, verbal backchanneling and emotional backchanneling. Nonverbal backchanneling is here characterized as all backchanneling feedback that is not verbal such as facial expressions (e.g. smile, frown, brow raise) and body movements (e.g. leaning, head movement). Verbal backchanneling feedback is all verbal given backchanneling behavior such as "yeah", "yes", "ok". Emotional backchanneling is backchanneling behavior that can be linked to text fragments that evoke a more emotional responses. In addition, for each dilemma the total amount of backchanneling per participant was calculated. Coinciding backchanneling behaviors, such as a "mmh" and a coinciding nod, were counted as one instance of backchanneling in the total amount of backchanneling.

The results suggest differences between people, conditions and also dilemmas. The backchanneling cue condition seems to influence the amount of backchanneling behavior more than the dilemma. See Figure 3.6 for the results of the different backchanneling cues.

After the experiment, the participants gave as feedback that the pauses were too long to appear natural. Additionally, it seemed as if they often used verbal backchanneling behavior to get the robot to go on with the dilemma during a pause instead of using it to indicate understanding. Furthermore, the amount of backchanneling cue moments were too many to feel natural.

3.6 Final Experiment

The task used during the third experiment was the same as during the second pilot experiment (see section 3.5). The text of the dilemmas told during the experiments were changed to make them easier to understand and to make them more human-like. Furthermore, the amount of possible backchanneling moments were decreased (see Figure 3.7). During the pilot experiment possible backchanneling moments were added after every sentence.

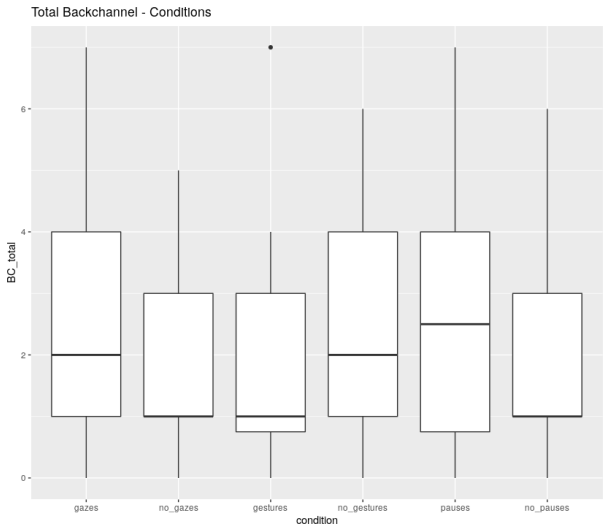


Figure 3.6: Amount of backchanneling during the dilemmas for the different backchannel cues during the second pilot experiment.

Robot: Okay, so you are standing next to some railway tracks. And you are seeing a trolley speeding towards you.; And further down the tracks, you see five people that are tied up and not able to move off the tracks.; Next to you is a lever, and you know that if you pull it the trolley would change tracks.; So you look over to the other track, and there you see another tied up person.; So if you pull the lever, the trolley would change tracks and the five people would be saved but the person tied up on the other tracks would die. And if you would do nothing, the trolley would continue on its way and the five people further down the track would die.
So what do you do?

Figure 3.7: This figure shows one of the nine stories that the robot narrated to each participant during the experiment. The red semicolon shows when the robot exhibited one of the robot cue conditions C1-C8. The cue condition was chosen randomly before narrating a story.

Since the participants gave the feedback that there were too many possible backchanneling moments, the possible backchannelling moments were limited to moments when an idea was finished.

3.6.1 Experimental Setup

The experiment was conducted in an office at the department of Computing Science at the University of Umeå in Sweden. Before the experiment started, the participant was seated at a table and given an information sheet about the experiment together with a consent form. Additionally, the experiment was explained by the test leader to ensure that no questions about the experiment remained. The participants were allowed to end the experiment at any time, and the data would then not be used. After signing the consent form, the participant was then seated in front of the robot at approximately 70 cm away. The experiment was video taped by a video camera situated behind the robot. The test leader started the video recording, left the room, and started the experiment

Robot: Recently I have been interested in ethical dilemmas.; I would like to talk with you about some of them and how you would respond. So I will tell you eight dilemmas and Adna will tell you one; After each dilemma I will ask you how you would respond. Then when you are finished, you will be asked to answer some short questions on the tablet. And afterwards I will explain the next dilemma to you. ; Ok, let's start with the first dilemma!

Figure 3.8: This figure shows the text used by the robot to explain the experiment. The red semicolon shows when the robot exhibited one of the robot cue conditions C1-C8. The cue condition was chosen randomly before narrating a text.

Robot: Thank you for participating in the experiment!; It was great hearing your responses to the dilemmas, they really gave me a new insight.; Adna will give you a questionnaire now, to gather some additional information.; I hope to see you again very soon!

Figure 3.9: This figure shows the text used by the robot after the last dilemma to thank the participant for the participation. The red semicolon shows when the robot exhibited one of the robot cue conditions C1-C8. The cue condition was chosen randomly before narrating a text.

from outside the room. The test leader was able to hear and see what was happening in the room using a video connection to a laptop situated on the desk in the test room.

The robot started by explaining the experiment using a random cue condition (see Figure 3.8). The first dilemma was randomly told by the test leader or the robot. If the first dilemma was told by the robot, then the test leader would explain the last dilemma. The other eight dilemmas were told by the robot. During each experiment, each one of the eight conditions was used once. The order of the dilemmas and the used cue condition were assigned randomly before the start of the experiment. After each dilemma, the robot asked the participant how they would react to the described dilemma. The participant answered verbally. The end of the answer by the participant was determined by the test leader, the robot then reacted with one of four random responses to show interest in the answer, thereby engaging the participant more. The responses were "Interesting", "Ok, thanks for your response", "Interesting response", and "Mmh, interesting". Next, the participant was asked to answer four questions about the conversation on the tablet (see Figure 3.3 and section B.2). The order of the questions was randomized to ensure that the participants would read the question each time. The robot ended the experiment with a short stop text after the nine dilemmas (see Figure 3.9). Afterwards, the participant was asked to fill out a post-questionnaire about their cultural background, familiarity with robots, and attitude towards the robot (see Appendix C). The questionnaire on attitude towards the robot was based on the Godspeed questionnaire by Bartneck et al. [2]. The first five questions can be characterized as being about the anthropomorphism, questions six to ten about animacy, questions eleven to fifteen about likability and the last three questions about the perceived safety. Afterwards, the participants were debriefed about the experiment by the test leader. The participants were rewarded with a sandwich and a drink after completing the experiment.

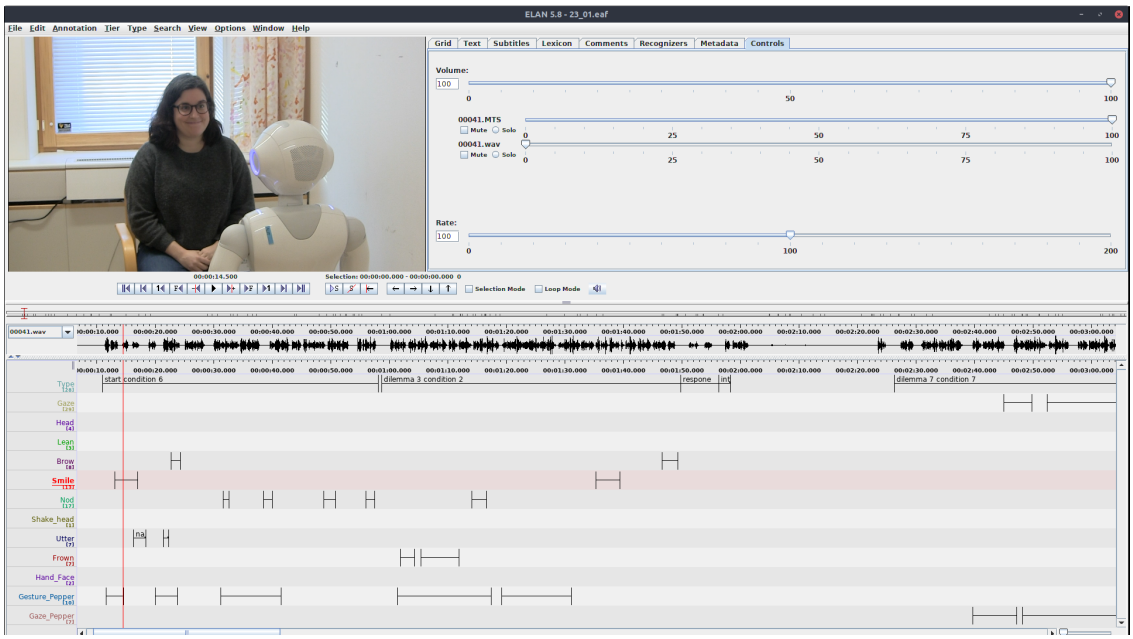


Figure 3.10: View of ELAN in the annotation mode, showing the video in the upper left, the audio stream in the middle and the annotations in the lower part of the image.

3.6.2 Video Analysis

The interaction was video recorded such that the backchannel behavior could be analyzed in detail afterwards. Video recordings are used in backchanneling research when not only verbal but also non-verbal backchannels are investigated. The videos were analyzed for backchanneling behavior by the participant (see Table 3.2) and backchannel cues by the robot using the video-annotation software ELAN [30]. In total 27 videos with an average length of 15 minutes were analyzed. First the start- and endpoints of each dilemma, response of the participant (i.e. how they answered the question about how to react during to the dilemma), and response of the robot (i.e. how the robot reacted to the response of the participant) were marked. Next, the backchannel cues by the robot were annotated with start- and endpoints. Then, the start- and endpoints of the backchannel behaviors of the participant were marked and if necessary transcribed according to backchanneling behavior used by the test participants. Examples of backchannel feedback that had to be transcribed are *utterances* (what is being said) and *head movement* (which direction). An example of how the data looks in ELAN after annotation and transcribing can be seen in Figure 3.10. Finally, all annotations were exported into a CVS file for further analysis.

3.6.3 Participants

In total 30 participants (21 male, 9 female) participated in the experiment. They were recruited among employees and students at the university. All participants had either a university degree or were university students. The age ranged from 19 to 56 years, with an average of 30 years. Most participants worked in computer science or related fields, except for two participants who worked with linguistics and law. The cultural background of the participants was diverse.

Two participants chose to interrupt the experiment, and for one participant, the experiment had to be terminated due to technical problems with the robot. One participant

was removed due to technical problems during the experiment.

Chapter 4

Results

In this section, the results of the video analysis of the experiment described in section 3.4 are presented. The presented results include the amount of backchanneling feedback per backchanneling cue condition, the results of the feedback questions, post-questionnaire, differences in backchanneling to the robot or a human, and cultural backchannel differences.

4.1 Backchanneling per Condition

The amount of backchanneling varies from person to person (see section 2.2). Therefore, the backchanneling scores for each participant were normalized between 0 and 1 using Equation 4.1 to remove individual differences.

$$z = \frac{x - min(X)}{max(X) - min(X)} \tag{4.1}$$

where x the score that should be normalized and X is the vector of all scores of one participant. When normalizing the scores for each participant and condition, then x is the total number of backchanneling feedback per condition and participant.

The smallest averaged scores of backchanneling responses was found for C1, i.e. when no backchanneling cue is used by the robot, see Figure 4.1. The largest average score of backchanneling responses was found for C8, which is the condition in which all cue conditions are present.

The data was analyzed using a linear mixed effect model. The fixed effects were gaze, gesture, and pause (present vs. not present) and the random factors were participant, dilemma, sex, experiment number, length of experiment, and responses to the post-questionnaire. AIC values were compared to determine which model fit the data best, with a complex model being preferred over a simpler model only if its AIC value is two or more points lower. The final model is presented in Appendix D Table D.1 . The best model consists of pause as fixed factor and participant as random factor. The influence of pauses on the amount of backchanneling is significant at a $p<0.1$ level.

When comparing the normalized scores for the different backchanneling cues, pauses seem to have the biggest influence on the amount of backchanneling, see Figure 4.2. The differences between the other backchanneling cues are very small. *Pauses* seem to have a positive effect on the amount of backchanneling. The difference between *gaze* being present or not is very small and no real effect on backchanneling can be seen.

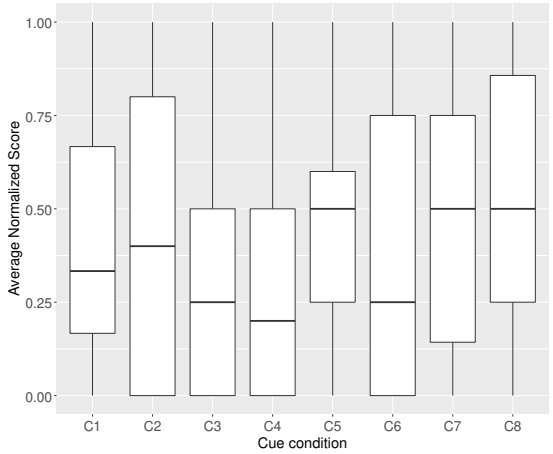


Figure 4.1: Boxplot of normalized scores of backchanneling responses for each backchanneling condition.

Cue condition	Mean	SD
C1: None	0.42	0.36
C2: Gesture	0.47	0.40
C3: Gaze	0.33	0.33
C4: Gesture+Gaze	0.32	0.36
C5: Pause	0.45	0.32
C6: Gesture+Pause	0.41	0.38
C7: Gaze+Pause	0.46	0.38
C8: Gaze+Gesture+Pause	0.50	0.37

Table 4.1: Normalized scores of the amount of backchanneling responses, for each one of the eight cue conditions.

Condition	Lean	Brow Raise	Smile	Frown	Nod	Shake Head	Head Move	Utter
C1: None	9.6	11.5	7.7	28.8	21.2	0.0	21.2	0.0
C2: Gesture	7.8	14.1	6.3	20.3	29.7	0.0	21.9	0.0
C3: Gaze	8.2	18.4	10.2	26.5	10.2	0.0	26.5	0.0
C4: Gesture+Gaze	7.8	11.8	23.5	9.8	27.5	2.0	13.7	3.9
C5: Pause	6.7	8.3	15.0	18.3	25.0	0.0	21.7	5.0
C6: Gesture+Pause	7.1	8.9	10.7	14.3	26.8	0.0	23.2	8.9
C7: Gaze+Pause	8.6	10.3	12.1	15.5	32.8	1.7	29.0	0.0
C8: Gaze+Gesture+Pause	7.2	11.6	11.6	14.5	29.0	1.4	21.7	2.9
Average	7.8	11.8	11.8	18.3	25.7	0.7	21.1	2.6

Table 4.2: Percentage of backchanneling behaviors used by the participants for each cue condition.

The percentages of backchanneling behavior used during each cue condition were calculated by summing up all instances of backchanneling feedback for a certain cue condition and backchanneling behavior and dividing by the total amount of backchanneling feedback of each cue condition, see Table 4.2. For these results one participant was excluded who backchanneld on average more than twice as often as the other participants which would have highly influenced the results. The most used backchanneling behavior in all conditions is *nodding*, followed by *head movement*. The least used backchanneling behavior is *shaking the head*, followed by *utterances*, these two backchanneling behaviors are also not present during all cue conditions. There are some differences in these orderings for the individual cue conditions. When the robot uses no cues (C1), the most used feedback is *frowning*, whereas *nodding* and *head movements* are used second most. The most used backchannel behavior, when only gaze is present as cue condition (C3), is *head movement* followed by *frowning*. *Nodding* is used a lot less than during the other cue conditions and is the third least used backchanneling behavior for cue condition three. When gestures and gaze is present as cues (C4), the participants *smile* twice as often than on average. The most *utterances* were used when the robot used gaze and pauses (C7).

A Pearson's Chi-squared test did not show a significant (p=0.24) relationship between the backchannel behavior and cue condition.

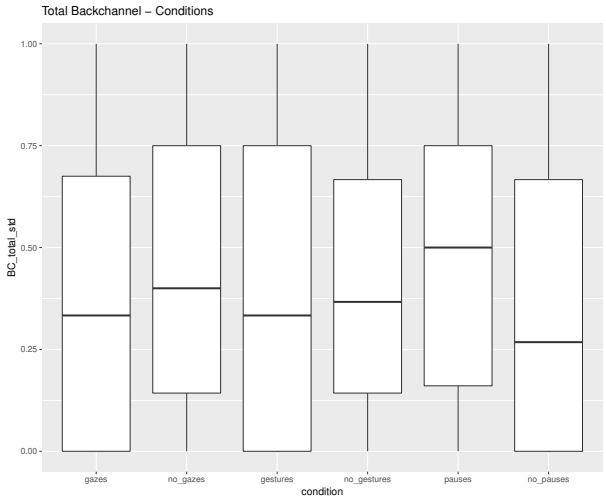


Figure 4.2: Boxplot of normalized number of backchannel responses for the three backchannel cues.

	Overall score		Enjoyment		Understandability		Engagement	
Condition	Mean	SD	Mean	SD	Mean	SD	Mean	SD
C1: None	3.44	1.00	3.28	1.14	4.00	1.12	3.80	1.04
C2: Gesture	3.28	1.14	3.28	1.06	4.00	1.12	3.76	1.05
C3: Gaze	3.36	1.08	3.36	1.04	4.24	0.93	3.84	0.94
C4: Gesture+Gaze	3.36	0.91	3.36	0.95	3.64	1.35	3.92	0.81
C5: Pause	3.16	0.90	3.16	1.07	3.76	1.30	3.88	1.01
C6: Gesture+Pause	3.32	0.95	3.28	0.89	3.96	0.98	3.76	0.97
C7: Gaze+Pause	3.48	0.82	3.56	1.04	3.84	1.11	3.76	1.05
C8: Gesture+Gaze+Pause	3.32	1.07	3.40	1.08	4.08	1.19	3.92	1.00
Average	3.34	0.97	3.34	1.02	3.94	1.14	3.83	0.97

Table 4.3: User ratings for the four feedback questions asked after each dilemma for each cue condition.

4.2 Feedback Questions

The responses to the feedback questions is on average relatively similar over all conditions, which means that the conditions did not have a great effect on the self-reported answers to the questions, see Table 4.3. All variables were on average rated higher than three. *Understandability* has been rated highest followed by *engagement*. The variability of how the participants scored the *understandability* is also the biggest, with the condition *gesture* and *gaze* (condition 4) being rated lowest (3.64) and the condition with only *gaze* (condition 3) being rated highest (4.24).

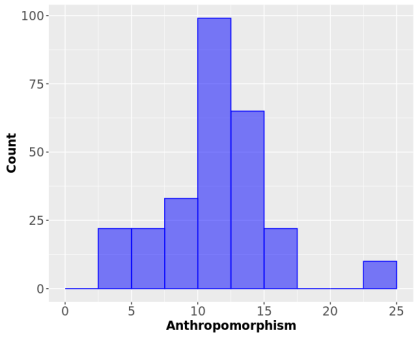
4.3 Post-Questionnaire

4.3.1 Technological Experience

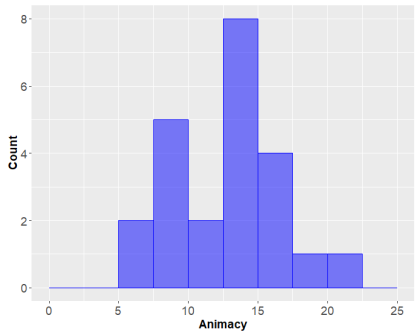
The majority of participants did not have a robot at home. Only five participants answered that they had a robot at home, and most of these robots were vacuum cleaners. Five of the participants had interacted with Pepper before, either in an other experiment, or out of curiosity. Most of the participants had not interacted with a different social

	Mean	SD	Min	Max
Anthropomorphism	11.79	3.84	5	23
Animacy	12.75	3.82	6	21
Likability	19	3.39	13	25
Emotional State	12.12	1.82	8	15

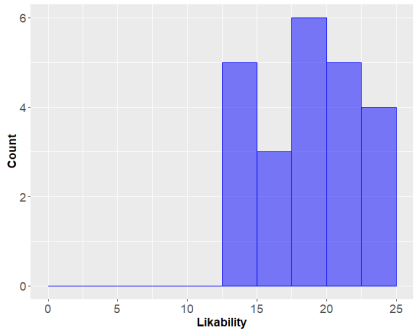
Table 4.4: Scores on the post-questionnaire. The maximally possible score on the anthropomorphism, animacy and likability questions is *25* and the maximally possible score on the emotional state questions is *15*.



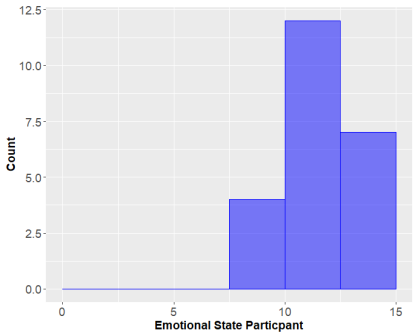
(a) Histogram for the five anthropomorphism questions.



(b) Histogram for the five animacy questions.



(c) Histogram for the five likability questions.



(d) Histogram for the three questions about the emotional state of the participant.

Figure 4.3: Histograms of the scores on the post-questionnaire.

robot than Pepper before the experiment. Only five partcipants had prior experience with other social robots, mostly with the Nao robot while participating in a different experiment. Most of the participants did not use a voice-controlled virtual assistant on their phone, two answered that they use them rarely, five sometimes and five almost daily. Six participants answered to use a voice-controlled virtual assistant using a smart speaker, one of them rarely, one sometimes and four almost daily.

4.3.2 Attitude Towards Pepper

The scores for anthropomorphism and animacy have a big variability with the minimum score for anthropomorphism being 5 and the maximum 23 out of 25 points, and the minimum score for animacy being 6 and the maximum 21 out of 25 points, see Table 4.4. The average score per question is 2.39 for anthropomorphism, 2.55 for animacy and 3.8 for likability. The emotional state of the participants was on average quite high with an average of 12.12 out of 15 possible points, see Table 4.4. The average score per question is

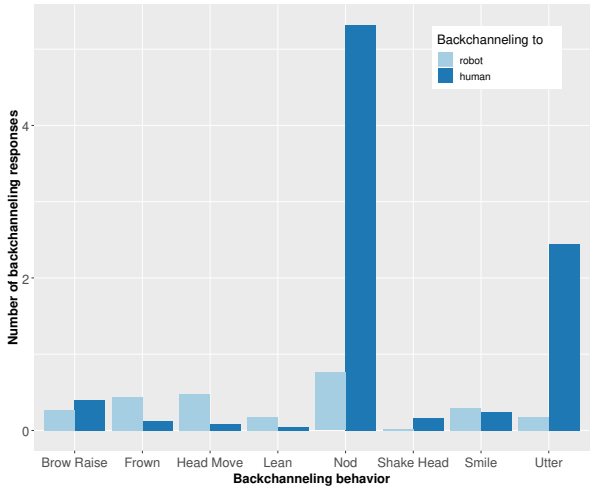


Figure 4.4: Number of backchanneling responses for the eight observed human backchanneling behaviors, with separate bars for backchanneling to the robot and to the human testleader respectively. The height of each bar is the average over all stories and participants.

Backchanneling Response	Robot	Human
Lean	0.19	0.04
Brow Raise	0.28	0.42
Smile	0.29	0.25
Frown	0.44	0.13
Nod	0.61	5.38
Shake Head	0.02	0.17
Head Movement	0.50	0.08
Utter	0.06	2.42

Table 4.5: Number of backchanneling responses observed while the participant is listening to the robot or testleader. The numbers are averaged over all stories and participants.

Backchanneling Response	Robot	Human
Lean	7.84%	0.47%
Brow Raise	11.77%	4.70%
Smile	11.98%	2.82%
Frown	18.30%	1.41%
Nod	25.71%	60.56%
Shake Head	0.65%	1.88%
Head Movement	21.13%	0.94%
Utter	2.61%	27.23%

Table 4.6: Percentage of backchanneling responses observed while the participant is listening to the robot or testleader.

4.04. The two lowest scores were explained by the participant by not feeling uncomfortable because of the robot but because of the dilemmas.

4.4 Differences in Backchanneling to a Robot and a Human

When looking at the differences in how the participants backchanneld to the robot and the test leader, it should be noted that each participant only had one trial with the test leader and eight with the robot. The results for how the participants backchanneled to the test leader are thus less reliable.

4.4.1 Backchanneling Behavior

For backchanneling to the robot, the most used backchanneling behaviors were *nodding*, followed by *head movement*, and *frowning*. For backchanneling to the human, the most used backchanneling behaviors were *nodding* and *utterances*. The distribution of used backchanneling behaviors is less spread out when backchanneling to the human compared to the robot. When comparing the percentages of used backchanneling behavior, see Figure 4.5, the second and third most used backchanneling behaviors when listening to the robot were only used rarely in instances that the participant backchanneled to the human.

Overall the participants backchanneled less to the robot than to the human, see Fig-

	Overall score		Enjoyment		Understandability		Engagement	
Condition	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Human	3.96	0.93	3.74	1.21	4.39	0.72	4.22	0.74
Robot	3.30	0.97	3.31	1.03	3.90	1.14	3.82	0.99
t-test	$p < 0.05$		$p \approx 0.1$		$p < 0.01$		$p < 0.05$	

Table 4.7: Comparison of the user ratings for the four feedback questions asked after each dilemma.

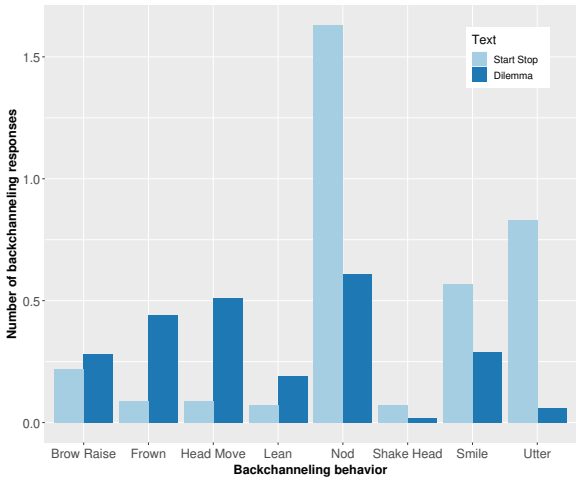


Figure 4.5: Number of backchanneling responses for the eight observed human backchanneling behaviors, with separate bars for backchanneling while listening to a dilemma and the start or stop text respectively. The height of each bar is the average over all stories and participants.

Backchanneling Response	Dilemma	Start-Stop
Lean	0.19	0.07
Brow Raise	0.28	0.22
Smile	0.29	0.56
Frown	0.44	0.09
Nod	0.61	1.63
Shake Head	0.02	0.07
Head Movement	0.50	0.09
Utter	0.06	0.83

Table 4.8: Number of backchanneling responses observed while the participant is listening to the robot telling a dilemma or start or stop text. The numbers are averaged over all stories and participants.

Backchanneling Response	Dilemma	Start-Stop
Lean	7.84%	1.84%
Brow Raise	11.77%	6.13%
Smile	11.98%	15.95%
Frown	18.30%	2.45%
Nod	25.71%	46.01%
Shake Head	0.65%	1.84%
Head Movement	21.13%	2.45%
Utter	2.31%	23.31%

Table 4.9: Percentage of backchanneling responses observed while the participant is listening to the robot telling a dilemma or start or stop text.

ure 4.4. Summed over all backchanneling behaviors, the average number of backchanneling responses per story and participant to the robot was 2.39, and to the human 8.88. Hence the participants backchanneled 3.72 times more often to the human.

4.4.2 Feedback Questions

The participants scored the feedback questions higher, i.e. found the conversations better, when the dilemmas were told by the test leader and not by the robot, see Table 4.7. The biggest difference between means can be found for the *overall score* with a difference of 0.62. T-test between the different user rating for the dilemmas told by the test leader and the robot were significant($p<0.05$), except for enjoyment which was only modestly significant ($p\approx 0.1$).

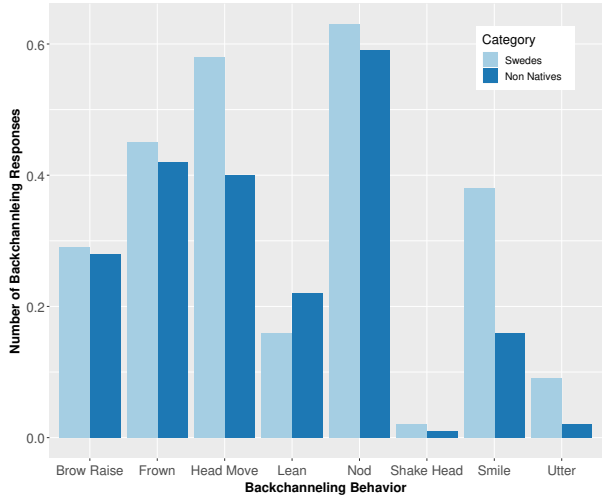


Figure 4.6: Number of backchanneling responses for the eight observed human backchanneling behaviors, with separate bars for backchanneling by native Swedes and non-Swedes respectively. The height of each bar is the average over all stories and participants.

4.5 Difference Between Dilemmas and Start-Stop Text

The most used backchanneling behaviors when listening to the start or the stop text was *nodding*, followed by *utterances*, and *smiling*, see Figure 4.9. The backchannel response *nodding* was used in almost 50% of the time. In contrast to this, *utterances* were one of the least used backchanneling behavior while listening to a dilemma. The backchanneling behavior *frowning* was often used by the participants when listening to the dilemma but not when listening to the start or stop text. Furthermore, the often used backchannel response *head movement* when listening to the dilemmas is almost not used when listening to the start-stop text.

Overall, the participants did backchannel a bit more while listening to the start or stop text in comparison to a dilemma, see Figure 4.8 and Figure 4.5.

4.6 Cultural Backchanneling Differences

The majority of the participants (17) reported that their native language was Swedish. Other reported native languages were: English (2), Danish(2), Spanish (2), Finish (1), German (1), Italian (1), and Portuguese (1).

Twelve participants reported that they lived in a different country than Sweden the last five years. They reported to have lived in the United Kingdom (4), Italy (3), France(2), Finland (2), Spain (2), Australia (1), Brazil (1), Iran (1), Germany (1), and the USA (1). Some of the participants reported that they lived in more than one country other than Sweden in the last five years.

On an English proficiency scale with the levels *basic*, *intermediate* , *advanced* , *high* and *native speaker*, the self-reported levels were advanced (9), high (16), and native (3).

As the only two big groups were native Swedes and non-native Swedes (i.e. not raised in Sweden), we only looked at the differences between these two groups as the other groups were too small to draw any conclusions from.

4.6.1 Differences Between Native Swedish and Non-Swedish

Native Swedish participants backchanneled on average 36.6 times during the whole experiment, to either the robot or the test leader. Non-Swedish participants backchanneled on average 32.2 times during the whole experiment. When the robot told the dilemmas, the Swedish participants backchanneld on average 2.6 times and the non-Swedish participants 2.1 times per dilemma. During the start or stop condition, the native participants backchanneld on average 0.6 times and the non-Swedes 0.8 times. The native Swedish participants backchanneled on average 8.2 times when listening to the test leader and the non-Swedes 9.8 times.

The used backchannel behavior by the native Swedes were not significantly different from the non-Swedes, and only small differences can be observed, see Figure 4.6.

Chapter 5

Discussion

This chapter provides answers to the three research questions, stated in section 1.3. The chapter is finalized by discussing limitations of the study.

5.1 Research Question 1

How Should a Robot Communicate in Order to Trigger a Human's Natural Backchanneling Behavior?

When examining the effect of the backchanneling cue condition, there are two aspects that we want to look into closer. First, do the backchanneling cues by the robot affect the backchanneling behavior of the human listener. Second, what is the effect of the individual backchanneling cues and is there a joint effect of the backchanneling cues, i.e. if the effect of multiple backchanneling cues at the same time is bigger than the individual effects.

The average amount of backchanneling behavior by the human listener is the lowest when no backchanneling cues are present during the telling of the dilemma by the robot (C1). This suggests that the use of backchannel cues has a positive effect on how often the human listener uses backchannel feedback.

The backchanneling cue *pause* was found to have positive effect on the amount of backchanneling feedback given by the listener, whereas *gesture* and *gaze* do not show a big influence on the backchanneling behavior. Truong et al. [28] found that there was visual backchanneling often coincided with mutual gaze. The lack of backchanneling feedback when the condition *gaze* was present could be explained by the lack of mutual gaze and thus less visual backchannel behaviors will be exhibited by the listener. Furthermore, Kendon [19] found that people look less at the other when the other is looking less at them. This could have provoked the listener to look less at the robot and thus also give less feedback. The lack of effect of the condition *gesture* could be allocated to the fact that the gestures were chosen at random and not according to the content of the dilemma.

A joint effect of the backchanneling cues seem probable given the fact that C4 to C8 have the highest average amount of backchanneling, except for C2. C2 seems to be an outlier as the high average is not visible for the cue condition *gesture* in combination with one of the other two cue combinations.

In conclusion, the results of the experiments indicate that the most influencing backchanneling cue of the robot is *pausing* and the influence of the backchanneling cue is even bigger when combined with other backchanneling cues. Furthermore, when the robot is looking less at the human (backchanneling cue *gaze* present) then the listener will backchannel

less. Thus to trigger more backchanneling behavior by the listener, the robot should look at the human.

5.2 Research Question 2

Does Such Human Backchanneling Affect How the Human Listener Is Perceiving the Interaction?

All four feedback questions got on average more than three out of the five possible points, which means that the participants found the conversations enjoyable, easy to understand and engaging and gave the conversations a good score. The cue condition did no have a big influence on how the conversations were scored on the different scales. The biggest difference in scoring can be seen for *understandability* where the cue condition *gaze* was scored highest and *gesture and gaze* lowest, but this effect cannot be seen when *pauses* were also present. When *pauses* were present, *gesture and gaze* scored better than only *gaze*. This indicates a negative influence of *pauses*, but this influence is also not clearly visible in the data.

Most of the participants had not interacted with a social robot before. This was probably good for the experiment since people who interacted with Pepper or other social robots before probably knew about the restrictions of such robots. People often expect more from robots than they are actual capable of doing Jung et al. [17] found that workers working together with robots often already use backchanneling even though the robot is not capable of understanding backchannel feedback. This naive judgment of the robots capabilities may have led to a more genuine and natural backchanneling behavior towards the robot.

The robot was on average scored as not human-like, and in between sentient and non-sentient. The reason could have been that the robot did not react to the feedback given by the human and also could not repeat any questions or dilemmas. The fact that the robot was scored as not human-like and non-sentient could have influenced how much the participants backchanneled to the robot when they realized that the robot did not react to their feedback.

The participants scored the robot relatively high on the likability scale. This could have made the participants communication more natural, and relaxed. The feeling of being relaxed was also shown in the response to the emotional state questions. Overall the participants felt comfortable, relaxed, and calm.

In summary, the results from the experiment could not show that the backchanneling cues had any effect on how a listener is perceiving the conversation with the robot.

5.3 Research Question 3

How Does a Humans Backchanneling Behavior Differ When Listening to a Robot Compared to a Human?

The most striking difference when comparing how the participants backchanneled to the robot and the test leader[1], was the amount of backchanneling feedback per dilemma. The participants backchannel 3.7 times more when listening to the test leader than when listening to the robot. One explanation for this difference could be that the participants do not expect the robot to react to the feedback given by them and thus backchannel

[1]The test leader during the experiments was Adna Bliek

less. This would also be in line with the fact that the participants scored the robot low on how responsive it was, and scored it more apathetic than reactive. On the other hand, most participants did not have prior experience of social robots and did thus not know whether the robot was capable of using the feedback or not. Furthermore, as stated before, people do use backchanneling in conversations with robots even though they are not equipped to process this information. But it should be noted that Jung et al. [17] did not report whether the actions used by the human towards the robot were as frequent as when collaborating with a human. Another explanation for the difference in backchannel frequency could be that the participants looked less at the robot than at the test leader. When the robot told the dilemma, some of the participants looked at the robot the whole time, others looked at the robot sometimes and away at other times, while others looked away from the robot most of the time. When listening to the test leader most participants looked at the test leader most of the time, looking away occasionally. The difference in how much the participants looked at the robot and the test leader could have influenced the amount of backchanneling.

Another interesting difference between the way the participants backchanneld to the robot and test leader is which backchannel responses was used. When giving backchannel feedback to the test leader, nods were used in more than 60% of the cases. The corresponding number for giving backchannel feedback to the robot was 25%. When backchanneling to the test leader mostly two backchanneling behaviors were used (nods and utterances), while when backchanneling to the robot the participants used a wider range of different behaviors.

It seems like the participants used more *nods* and *utterances* but there is not a big difference between the other backchannel responses. *Nodding* can be characterized as a generic backchannel behavior, whereas other backchannel behaviors, for example *frowning*, can be categorized as specific backchannel behavior. It seems as if the amount of generic backchannel behavior increased more than the specific feedback comparing the backchanneling to the human and robot. *Frowning* is even more prevalent when backchanneling to the robot. One explanation for this difference could be that the listener uses generic backchanneling feedback to let the speaker know that they are still listening and that they do not feel the need to let the robot know that they are still listening as the robot does not use this feedback. An explanation for the increase use of generic backchanneling behavior could be that the listener feels more open to show emotions related to the dilemma in front of the robot, since the robot will not judge them for their feelings.

The difference between the behavior when the robot tells the dilemmas, and the start and stop text is not whether the participant is listening to a robot or a human, but lies in the nature of the text. While listening to a dilemma, the participant has to create a mental model of the situation, and think about how they would react. In contrast, the start and stop text give an introduction or end to the experiment. The participant already had the possibility to ask all questions about the experiment to the test leader beforehand and read the description of the experiment. The participant thus did not learn new information about the experiment during the start text. In addition to the difference in nature of the text, the length of the start and stop text was also shorter than the dilemmas.

The participants backchanneled during the start and stop text a bit more closely to how they backchannel while listening to the test leader. They used *nods* most often followed by *utterances*. A difference to how the participant backchanneled during the dilemmas to both the robot and the test leader can be seen in how often the participant

was *smiling*. The participant smiled approximately twice as often during the start and stop text than during the dilemmas. Firstly, this difference could be explained during the start text due to the novelty of the situation. The participants seemed amused and fascinated by the robot and showed this during the start text by smiling. Secondly, some of the participants were happy when the experiment was over due to its length and sometimes difficult questions, this prompted them to smile when the stop text was uttered by the robot. Another explanation could be that the text is more personal, the robots talks to the participant directly asking for their name and thanking them at the end of the experiment. This more personal text could have led to more smiling by the participants. In contrast, one does normally not smile when hearing ethical dilemmas where it is hard to make a decision.

Another difference that can be seen is that the participants almost never *frowned* during the start and stop text, whereas *frowning* was the third most often used backchannel behavior to the robot during the dilemmas. This difference could be again be explained by the nature of the texts. During the dilemmas emotional responses can be expected at certain moments which could result in *frowning*. These more negative emotional responses were not to be expected during the start or stop text.

The Swedish participants backchanneled overall a bit more than the non-Swedish participants. Additionally, the backchannel behavior used by the Swedish and non-Native Swedish participants did only differ slightly, these differences can probably be better explained by personal than cultural differences. The differences between the two groups of participants are not big enough to draw any conclusions. Fant [8] suggested that Scandinavian listeners use verbal-backchanneling to show attention. This use of verbal-backchanneling responses could not be seen by the Scandinavian listener., *Utterances* were even one of the least used backchanneling feedback. This lack of verbal-backchanneling could be attributed to the fact that the robot was not using this feedback to adjust its storytelling. The human storyteller in contrast to the robot could adjust its storytelling and here we see a rise in the amount of verbal feedback used by the participants.

To conclude, when comparing the backchanneling behavior when listening to the robot compared to the human, the listener backchanneled more when listening to the human and uses more *utterances*. Furthermore, the listener used more generic backchanneling responses. When the listener backchanneled to the robot, they backchanneled less often and seemed to show their surprise more openly. In addition, the backchanneling response seems also to be dependent on the context of the text they were listening to.

5.4 Limitations

This study was a relatively small study with only 27 participants that could be analyzed. The results should therefore only be seen as an indicative on how and why humans backchannel to robots, and how this affects the conversation.

Limitations of this study also include that the robot could not react to the backchannel feedback from the human listener, which could have reduced the amount of backchanneling feedback, as well as a differentiation between different backchanneling-inviting cues by the robot. Furthermore, this also makes it hard to compare the participants' backchanneling to the robot and to the human speaker.

Another limitation of the study that might have influenced the results and how the robot was perceived is that the robot was not able to repeat any dilemmas or questions. When participants needed clarification this had to be given by the test leader. This might

have reduced the illusion of the robot being able to react to the participant

Chapter 6

Conclusion

This research has only been a first investigation into how backchanneling cues by the robot affect human backchanneling behavior. The results show that humans backchannel naturally to a robot. Furthermore, it also suggests that the used backchanneling feedback is dependent on the task as well as whether the story is told by a robot or a human.

6.1 Future Research

To investigate the interaction further another study should be conducted with more participants. In a future study, the backchanneling cue *gaze* could be excluded as it did not seem to have a positive effect on the amount of backchanneling feedback. Furthermore, by removing one backchanneling cue, the other conditions could be tested multiple times for each participant which could give more insight into the backchanneling behavior of an individual and the influence of story and backchanneling cue without increasing the time needed for the experiment.

In a future research we would suggest making the robot able to repeat the key aspects of a story and the question posed after the story. This addition could make the robot seem more alive and responsive and would prevent the participant having to ask the test leader for clarification which interrupts the experiment. Another possibility to make the robot more reactive would be to use a Wizard of Oz experiment. During a Wizard of Oz experiment, the robot would be controlled by the test leader without the knowledge of the participant.

Another interesting investigation could be to repeat the experiment with another humanoid robot, e.g. the Nao robot. Repeating the experiment with another robot could show whether the found results are translatable to other robots.

Appendix A

Dilemmas

1. Okay, so you are standing next to some railway tracks. And you are seeing a trolley speeding towards you. And further down the tracks, you see five people that are tied up and not able to move off the tracks. Next to you is a lever, and you know that if you pull it the trolley would change tracks. So you look over to the other track, and there you see another tied up person. So if you pull the lever, the trolley would change tracks and the five people would be saved but the person tied up on the other tracks would die. And if you would do nothing, the trolley would continue on its way and the five people further down the track would die.

2. So, you are again standing next to railway tracks and again you see a trolley speeding towards you. You look down the track and see that there is one person tied up. And you recognize that person, they are a good friend of yours. Like last time, you stand next to a lever and you can pull it to change the track the trolley is on. But on the other track you see that five people are tied up, you look to see whether you know any of them but you don't. So if you do nothing the trolley would hit your friend and they would die. And if you would pull the lever, the trolley would change tracks, and the five people on the other track would die, but your friend would survive.

3. Ok, so this is another trolley problem. Unlike the previous times, you stand on a bridge above the tracks, and not next to them. You again see a trolley speeding towards you and you look further down the track, where you see five people tied up. You realize that you could stop the trolley if you put something heavy in front of it. The only two movable objects close by are you and a heavy stranger on the bridge. You make some quick calculations and you now know that the only way to stop the trolley would be by throwing the heavy stranger in front of it but you also know that this would kill them. You yourself are not as heavy as the stranger and could thus not stop the trolley. So, if you would push the stranger, the trolley would stop but the stranger would die. And if you do nothing the trolley would continue and the five people tied up would die.

4. You are a brilliant transplant surgeon. And you have five patients that are in critical condition, needing an organ transplant to survive. But unfortunately, there are no organs available to them. A young healthy traveler comes into your practice for a routine checkup. So you perform the checkup and they are totally healthy and you recognize that their organs would be a perfect match for all of your critical ill patients. You talk to the traveler and they say that they didn't tell anyone that

they were coming to you for the checkup. And so, if you would kill them, no one would suspect you. Thus, if you would kill the traveler, your five patients would survive. And if you would do nothing they would die.

5. Suppose that you are an inmate at a concentration camp together with your son. And your son has talked to you about fleeing, as he doesn't have any hopes of leaving the camp alive in a different manner. You remind him that the penalty for attempting to flee is death. Last night, your son tried to flee, but was stopped by the guards. Now, as the penalty for attempting to flee is death, a guard is about to hang your son. The guard tells you that you have to remove the chair from underneath your son. And if you don't do as asked he will not only kill your son but also another innocent inmate.

6. So you entered a cave close to the sea with a group of people. Now the water is rising again, as the high tide is coming up. One of the group members is a pregnant woman that is now leading your group on the way out of the cave. You all are almost outside when the pregnant woman gets stuck in the mouth of the cave. You try to move her in any direction, but no one in your group succeeds. You have to think of an idea quickly, as the cave will fill with water soon. A person in your group mentions that they have dynamite with them. This would be the only way to remove the pregnant woman, but it would also kill her. If you use the dynamite the pregnant woman would die but the rest of the group would survive. And if you don't use it, all group members would drown, except for the pregnant woman as her head is outside the cave.

7. Suppose that you work for a mining company. And one day, a group of miners gets trapped during work in a shaft. They could be in either one of two shafts and you don't have any way of communicating with them. While you are trying to figure out in which shaft the miners are, you are informed that floodwater is rising. Now, you have two possibilities, either block one of the two shafts or do nothing. If you block one shaft, the other one will fill completely with water and if the miners would be there they would all drown, but if they are in the shaft that you blocked, they would all survive. You only have enough sandbags to block one of the two shafts completely. You also have another option, which is to do nothing, in this case, one miner would drown no matter in which shaft they are.

8. So, a mad man has been arrested, after threatening to explode several bombs in crowded areas. But unfortunately, he has already planted the bombs and they will go off shortly. The authorities cannot make him tell them the location of the bombs by conventional methods. He refuses to say anything and requests a lawyer. As time is running out, a high-level official suggests to use torture, which would of course be illegal. But he thinks that it is ok to use torture in this situation since nothing else works and many people may die if they don't find the bombs.

9. Suppose that there are two married couples, Jane and Jasper, and Debbie and Dave. Jane and Debbie are both unhappy in their marriages and would like their husbands to be dead. So Jane acts on these feelings and puts poison in Jasper's coffee, which kills him. One day, Dave puts poison in his own coffee by accident, thinking it's cream. Debbie has the antidote to the poison but she doesn't give it to him. Debbie

knows that she would be the only one who could save Dave, but she lets him die. So, Jane actively killed her husband and Debbie did not help hers.

Appendix B

Feedback Questions

B.1 Map Task

1. How would you rate the overall conversation on a scale from 1(lowest) to 5(highest)?
2. How enjoyable did you find the conversation on scale from 1(not enjoyable at all) to 5(very enjoyable)?
3. How difficult or easy did you find it to follow the story on scale from 1(very difficult) to 5(very easy)?
4. How engaging did you find the story on scale from 1(not engaging at all) to 5(very engaging)?

B.2 Dilemmas

1. How would you rate the overall conversation on scale from 1(lowest) to 5(highest)?
2. How enjoyable did you find the conversation on scale from 1(not enjoyable at all) to 5(very enjoyable)?
3. How difficult or easy did you find it to follow the story on scale from 1(very difficult) to 5(very easy)?
4. How engaging did you find the story on scale from 1(not engaging at all) to 5(very engaging)?

Appendix C

Post-Questionnaire

C.1 Socio-Demographic Questions

1. What gender do you identify as?
   (a) Male
   (b) Female
   (c) Other:
   (d) Prefer not to answer
2. What is your age?
3. What are you studying/what is your profession or what are you working with?
4. In which country did you grow up?
5. What is your first language?
6. How long have you been living in Sweden?
7. Which language(s) do you speak most often?
8. Which languages are you capable of speaking? What is your proficiency level(Basic Knowledge, Intermediate Level, Advanced Level, High Level, Native Speaker)?
9. Have you been living in another country than Sweden in the last 5 years? If yes, which and for how long?

C.2 Technological Experience

1. Do you have a robot at home(e.g. vacuum cleaner robot, Sphero, Vector)? If yes, what kind of robot?
2. Did you interact with a Pepper robot before? If yes, for what purpose?
3. Did you interact with another social robot before? If yes, for with which robot and for what purpose?

4. Do you use voice-controlled digital assistants on your phone or computer(e.g. Siri, Google, Cortana)? If yes which one and how often?

5. Do you use voice-controlled digital assistants using smart speakers (e.g. Google Home, Alexa on the Amazon Echo, Apple HomePod)? If yes which one and how often?

C.3 Attitude Towards Pepper

Please rate your impression of Pepper on these scales:

Fake	1	2	3	4	5	Natural
Machine-like	1	2	3	4	5	Human-like
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly
Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive
Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Please rate your emotional state while interacting with Pepper on these scales:

Anxious	1	2	3	4	5	Relaxed
Agitated	1	2	3	4	5	Calm
Unpleasant	1	2	3	4	5	Comfortable

Appendix D

Statistical Tests

	Estimate	Standard Error	df	t-value	p
(Intercept)	2.4123	0.3914	30.8837	6.164	7.83e-07
pause	0.4018	0.2296	168.5681	1.750	0.082

Table D.1: Model = Amount of Backchanneling $\sim$ Pause + (1|Participant)

Bibliography

[1] Anne H Anderson et al. "The HCRC map task corpus". In: *Language and speech* 34.4 (1991), pp. 351–366.

[2] Christoph Bartneck et al. "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots". In: *International journal of social robotics* 1.1 (2009), pp. 71–81.

[3] Janet B. Bavelas, Linda Coates, and Trudy Johnson. "Listeners as co-narrators." In: *Journal of Personality and Social Psychology* 79.6 (2000), pp. 941–952. ISSN: 0022-3514. DOI: 10.1037//0022-3514.79.6.941. URL: http://web.uvic.ca/psyc/bavelas/2000listnrs.pdf.

[4] Janet Beavin Bavelas and Jennifer Gerwing. "The listener as addressee in face-to-face dialogue". In: *International Journal of Listening* 25.3 (2011), pp. 178–198. ISSN: 1932586X. DOI: 10.1080/10904018.2010.508675.

[5] Stefan Benus, Augustín Gravano, and Julia Hirschberg. "The Prosody of Backchannels in {A}merican {E}nglish". In: *Proceedings of the 16th International Congress of Phonetic Sciences* August (2007), pp. 1065–1068.

[6] Jasone Cenoz. "Pauses and hesitation phenomena in second language production". In: *ITL-International Journal of Applied Linguistics* 127.1 (2000), pp. 53–69.

[7] Allen T Dittmann. "Developmental factors in conversational behavior". In: *Journal of Communication* 22.4 (1972), pp. 404–423.

[8] Lars Fant. "Cultural mismatch in conversation: Spanish and Scandinavian communicative behaviour in negotiation settings". In: *HERMES-Journal of Language and Communication in Business* 3 (1989), pp. 247–265.

[9] Susan Goldin-Meadow and Cynthia Butcher. "Pointing toward two-word speech in young children". In: *Pointing: Where language, culture, and cognition meet* (2003), pp. 85–107. URL: https://cpb-us-w2.wpmucdn.com/voices.uchicago.edu/dist/c/1286/files/2019/04/KitaSotaro_2003_Chapter5PointingTowar_PointingWhereLanguage.pdf.

[10] Jonathan Gratch et al. "Creating rapport with virtual agents". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4722 LNCS (2007), pp. 125–138. ISSN: 03029743. DOI: 10.1007/978-3-540-74997-4_12.

[11] H Paul Grice. *Studies in the Way of Words*. Harvard University Press, 1989. DOI: 10.5040/9780571344444.ch-005.

[12] Bettina Heinz. "Backchannel responses as strategic responses in bilingual speakers' conversations". In: *Journal of Pragmatics* 35.7 (2003), pp. 1113–1142. ISSN: 03782166. DOI: 10.1016/S0378-2166(02)00190-X.

[13] Anna Hjalmarsson and Catharine Oertel. "Gaze direction as a backchannel inviting cue in dialogue". In: (2011), p. 1. URL: http://www.speech.kth.se/prod/publications/files/3785.pdf%7B%5C%%7D0Ahttp://wwwhome.cs.utwente.nl/%7B~%7Dkoki/rcva/hjalmarsson.pdf.

[14] Nusrah Hussain et al. "Speech driven backchannel generation using deep Q-network for enhancing engagement in human-robot interaction". In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH* 2019-Septe (2019), pp. 4445–4449. ISSN: 19909772. DOI: 10.21437/Interspeech.2019-2521. arXiv: 1908.01618.

[15] Benjamin Inden et al. "Micro-timing of backchannels in human-robot interaction". In: (2014).

[16] Benjamin Inden et al. "Timing and entrainment of multimodal backchanneling behavior for an embodied conversational agent". In: *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction* (2013), pp. 181–188. DOI: 10.1145/2522848.2522890.

[17] Malte F Jung et al. "Engaging robots: easing complex human-robot teamwork using backchanneling". In: *Proceedings of the 2013 conference on Computer supported cooperative work*. ACM. 2013, pp. 1555–1566.

[18] Daniel Jurafsky et al. "Lexical, prosodic, and syntactic cues for dialog acts". In: *ACL/COLING Workshop on Discourse Relations and Discourse Markers* (1998), pp. 114–120. URL: http://acl.ldc.upenn.edu/W/W98/W98-0319.pdf?origin=publication%7B%5C_%7Ddetail.

[19] Adam Kendon. "Some functions of gaze-direction in social interaction". In: *Acta Psychologica* 26.C (1967), pp. 22–63. ISSN: 00016918. DOI: 10.1016/0001-6918(67)90005-4.

[20] Sotaro Kita. *Cross-cultural variation of speech-accompanying gesture: A review*. Feb. 2009. DOI: 10.1080/01690960802586188. URL: http://www.tandfonline.com/doi/abs/10.1080/01690960802586188.

[21] Hanae Koiso et al. "An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs". In: *Language and speech* 41.3-4 (1998), pp. 295–321.

[22] Divesh Lala et al. "Detection of social signals for recognizing engagement in human-robot interaction". In: (2017). arXiv: 1709.10257. URL: http://arxiv.org/abs/1709.10257.

[23] Jin Joo Lee, Cynthia Breazeal, and David DeSteno. "Role of speaker cues in attention inference". In: *Frontiers Robotics AI* 4.OCT (2017), pp. 1–14. ISSN: 22969144. DOI: 10.3389/frobt.2017.00047.

[24] Han Z. Li, Yanping Cui, and Zhizhang Wang. "Backchannel Responses and Enjoyment of the Conversation: The More Does Not Necessarily Mean the Better". In: *International Journal of Psychological Studies* 2.1 (2010), pp. 25–37. ISSN: 1918-7211. DOI: 10.5539/ijps.v2n1p25.

[25] S. MAYNARD. "Analyzing interactional management in native/non-native English conversation : A case of listener response". In: *IRAL. International review of applied linguistics in language teaching* 35.1 (1997), pp. 37–60. ISSN: 0019-042X.

[26] David McNeill. *Hand and Mind: What Gestures Reveal About Thought*. 1994. DOI: 10.1177/002383099403700208. URL: http://www.cogsci.ucsd.edu/%7B~%7Dbkbergen/cogs200/McNeill%7B%5C_%7DCH3%7B%5C_%7DPS.pdf.

[27] L. Morency. "Modeling Human Communication Dynamics [Social Sciences]". In: *IEEE Signal Processing Magazine* 27.5 (2010), pp. 112–116.

[28] Khiet P Truong et al. "A multimodal analysis of vocal and visual backchannels in spontaneous dialogs". In: *Twelfth Annual Conference of the International Speech Communication Association*. 2011.

[29] Nigel Ward and Wataru Tsukahara. "Prosodic features which cue back-channel responses in English and Japanese". In: *Journal of pragmatics* 32.8 (2000), pp. 1177–1207.

[30] Peter Wittenburg et al. "ELAN: a professional framework for multimodality research". In: *5th International Conference on Language Resources and Evaluation (LREC 2006)*. 2006, pp. 1556–1559.

[31] Victor H. Yngve. "On getting a word in edgewise". In: *CLS-70*. University of Chicago, 1970, pp. 567–577.