# How does Modality Matter? Investigating the Synthesis and Effects of Multi-modal Robot Behavior on Social Intelligence

Karen Tatarian, Rebecca Stower, Damien Rudaz, Marine Chamoux, Arvid Kappas, Mohamed Chetouani

**Noname manuscript No.**
(will be inserted by the editor)

# How does modality matter? Investigating the synthesis and effects of multi-modal robot behavior on social intelligence

**Karen Tatarian · Rebecca Stower · Damien Rudaz · Marine Chamoux · Arvid Kappas · Mohamed Chetouani**

Received: date / Accepted: date

**Abstract** Multi-modal behavior for social robots is crucial for the robot's perceived social intelligence, ability to communicate nonverbally, and the extent to which the robot can be trusted. However, most of the research conducted so far has been with only one modality, thus there is still a lack of understanding of the effect of each modality when performed in a multi-modal interaction. This study presents a multi-modal interaction focusing on the following modalities: proxemics for social navigation, gaze mechanisms (for turn-taking floor-holding, turn-yielding and joint attention), kinesics (for symbolic, deictic, and beat gestures), and social dialogue. The multi-modal behaviors were evaluated through an experiment with 105 participants in a seven minute interaction to analyze the effects on perceived social intelligence through both objective and subjective measurements. The results show various insights of the effect of modalities in a multi-modal interaction onto several behavioral outcomes of the users, including taking physical suggestions, distances maintained during the interaction, wave gestures performed in greeting and closing, back-channeling, and how socially the robot is treated, while having no effect on self-disclosure and subjective liking.

K. Tatarian · D. Rudaz · M. Chamoux
SoftBank Robotics Europe, Paris, France
E-mail: name.surname@softbankrobotics.com

K. Tatarian · M. Chetouani
Sorbonne University, Institute for Intelligent Systems and Robotics, CNRS UMR 7222, Paris, France
E-mail: name.surname@sorbonne-universite.fr

R. Stower · A. Kappas
Jacobs University, Department of Psychology and Methods, Bremen, Germany
E-mail: firstinitial.surname@jacobs-university.de

**Keywords** Multi-modal interaction · Gestures · Gaze · Proxemics · Social Cues · Social Intelligence · Nonverbal Behavior · Human Robot Interaction

## 1 Introduction

By merely observing humans, one can directly infer that no social interaction takes place without cues, whether verbal or nonverbal, that allow others to interpret behaviors and reasonably estimate intentions [62]. Furthermore, those verbal and nonverbal cues have an effect on others by eliciting tangible change in their observable behavior or even internal changes, e.g., awareness of a particular social setting [76]. Moreover, proper communication and exchange of information is crucial to a human's need to feel connected, promote well-being, and gain acceptance by social groups [61]. However, these powerful social signals and nonverbal behaviors are complex and multi-modal. They are made of different combinations of modalities and cues such as *kinesics* (e.g., gestures) ([41], [23]), *gaze behavior* [36], and *proxemics* (e.g., management of space and environment) [31]. Similarly, these multi-modal nonverbal behaviors hold several functions, which include the ability to understand and manage others in social interactions and "act wisely in human relations", and as such contribute to one's *social intelligence*. [71].

In today's world, humans not only have to interact with each other, but also with machines, including robots. With robots gaining further presence in a human's everyday life, synthesizing and understanding these multi-modal behaviors is crucial to designing better and more appropriate human-robot interactions. In an attempt to solve this issue, some studies have been inspired by human-human interaction to design rule-

based algorithms targeted at investigating individual modalities ([6], [65], [47]). In contrast, other research has focused on data-driven learning methods designed to synthesize multi-modal behavior, however, lacking a clear understanding of the effect of each modality forming the multi-modal behavior ([43], [58]). Thus, there is still a need to investigate how such modalities can be combined and the effect and function of each when performed in a multi-modal combination. This will allow for a better understanding of how and when the robot could use combinations of different modalities to appear as a socially intelligent agent and express intentions and information using verbal and nonverbal behavior more naturally.

This paper presents a system of multi-modal behaviors comprised of the following modalities: gaze, kinesics, proxemics, and social dialogue. The system was evaluated objectively by studying the behavioral outcomes. In addition, critical to evaluating peoples interactions with social robots is also the extent to which they like the robot, or form a general positive impression during their interaction. Many studies in HRI capture liking directly, by asking questions such as 'I like [the robot]' ([72], [69], [52]), whereas others take a more indirect approach, assessing statements such as '[the robot] is kind / friendly / warm' ([45], [60]). Thus, in this study we also included a subjective measure of liking through a self-report "liking" scale [1]. Instead of analyzing the effects of each modality by contrasting them in isolation and thus losing possible coupling effects, this paper compares a version with all implemented modalities together with versions, in which each modality is subtracted in turn.

The remaining paper is organized as follows: Section 2 reviews background work done on each of the modalities to be integrated. In Section 3, the implemented system of multi-modal behavior is discussed. Section 4 presents the hypotheses as well as the design and set-up of the evaluation study. In Sections 5 and 6, the results are presented and analyzed and the findings are discussed in more depth. In Section 7 the paper is concluded.

## 2 Background

In the past decades, there has been considerable interest among psychologists and sociologists to investigate non-verbal behaviors observed in humans and used as communication methods and tools. Inspired by those findings, more recently, social roboticists have tried to synthesize these modalities on different robots in order to study their impact on human-robot interaction (HRI). These modalities include gaze, kinesics, and proxemics. This section highlights the human-human studies done in addition to the HRI research for each modality as well as social dialogue.

### 2.1 Gaze

In 1967, Kendon [36] was the first to classify and analyze gaze aversion in human-human interaction, claiming that humans in fact do not spend the majority of their time in a conversation directing gaze straight at another human's face. He concluded that gaze aversion was done for four primary reasons: turn-taking, turn-yielding, floor-holding, and intimacy regulation (used to regulate the level of shared emotional arousal) ([37], [75],[74]). Today, gaze mechanisms, including gaze aversion, are still a study of interest for social roboticists. Research in HRI has involved conducting studies to better understand social gaze ([53], [6]), using gaze to reference an object of conversation by joint attention ([40], [5]), designing gaze cues to modulate group conversation ([53],[54]), and regulating turn-taking in conversations ([6], [54]). In addition, conversational social gaze constructed of gaze aversions to perform role-signaling, turn-taking, and topic-signaling prompted high indices of likeability towards the robot [53]. Moreover, for robots which lack expressive eyes, head controlled tilts have been designed to convey gaze aversion [6]. The former concluded that while social gaze aversions did not increase the human's comfort in eliciting more self-disclosure, it did decrease interruption time caused by the user and the robot was perceived as more thoughtful. Additionally, the study analyzed the direction of gaze aversions in human-human interactions with respect to its three primary functions: cognitive, intimacy-regulation, and floor management [6].

An additional important function of gaze is joint attention. It supplies people with a way of interpreting and predicting each other's actions and focus attention [27]. For instance, speakers tend to use deictic expressions followed by a glance towards the object of reference [18]. Thus it is no surprise that joint attention attracted the attention of researchers in the HRI field. For instance, it was shown in [12] that users reached objects faster when they could follow the gaze of the robot iCub, who was giving instructions while glancing at referenced object. Similarly, [68] showed that users interacting with a robot that had a gaze with a reference function found it easier to complete a task than with a robot that had random gaze. Joint attention has proven to be functional for

[1] Further subjective measurements, referring to the comparison between self-reported attitudes and behaviors towards social robots will be examined elsewhere.

social robots to shift the human's attention to the spot at which it is looking [78]. In addition, for collaboration tasks involving object selection, robot gaze shifts assisting its speech were shown to be advantageous for cooperation specially when the human was required to choose the object being referred to by the robot as fast as possible ([2], [13]). Furthermore, in [51], during hand over tasks, users started reaching for an item much sooner when a robot consistently gazes at the handover area than when it gazes away from that point. In parallel, when gaze is used as part of a multi-modal behavior, it often has a supportive and enhancing role to other social behaviors, notably speech and gestures [3].

### 2.2 Kinesics

Gestures for humans have been categorized and defined primarily based on their role in communication and their functions as follows ([49],[41],[4],[23]):

- *Iconic gestures* for describing physical objects and events mentioned in a conversation; e.g., forming a small circle with the hand to refer to a small ball.
- *Metaphoric gestures* for depicting abstract concepts being referred to; e.g., fast back-forth hand movement to indicate 'ongoing' work.
- *Deictic gestures* for indicating objects in the physical space where the conversation is taking place; e.g., point at a road close by.
- *Emblem gestures* or symbolic gestures for expressing language-like features with agreed upon culturally specific properties; e.g., the V hand gesture with the index and middle fingers to indicate a peace sign.
- *Beat gestures* for emphasizing significant points or certain words in the speech using rhythmic movements of hands and arms e.g., hand gesture to indicate the introduction of a new topic

Gestures have also been studied and implemented on robots aiming to improve human-robot interaction. While deictic, beat, iconic, and metaphoric gestures were all found to boost the robot's performance as a narrator in a narrative scenario, deictic gesture significantly ameliorated the user's recall of information on the story [34]. Additionally, the robot which performed correctly timed nods in a conversation and proper gaze and gesture sharing behaviors was ranked more highly than a robot who did not have such behaviors [35]. Moreover, gestures play a role in portraying emotional expressions. For instance, submissiveness can be expressed by an open hand shape; on the other hand, dominance can be portrayed in a pointing hand shape [39]. Similarly in social robotics, modulating the robot's body movement by varying its head tilts and body expansiveness influenced perceived dominance [55].

Another important aspect of social interaction is *alignment*, which refers to the convergence of linguistic behavior and/or similarity in mental representation ([57],[14]). Alignment is an ubiquitous feature used to measure to which extent interactions shape behavior and their success at communicating shared understanding [33]. For instance, it was shown that alignment, through mimicry of postures, mannerisms, and facial expressions in dyadic interactions (chameleon effect [16]), increased the rapport between the participants, the pro-social behavior even beyond the interaction and smoothness of interaction [9]. Moreover, a study found that users, when retelling a story to a third participant, were more likely to demonstrate the same iconic gestures they witnessed the first time [50]. Alignment equally plays an important role in human-computer interaction in enhancing communicative success [14]. In robotics, people have been found to nod more when interacting with a robot that nods along in response to that in comparison to a robot who does not mirror their nodding [66]. In addition, a computational method for evaluating and modeling of interpersonal synchrony in behaviors during interactions offered a perspective for building social interfaces for robots and embodied conversational agents [20]. Furthermore, motor resonance, which is the activation of the observer's motor control system during action perception, was used to not only produce more natural interactions for robots with humans but also as an evaluation method to determine quantitatively how the robot is perceived by the human [64].

### 2.3 Proxemics

Proxemics, which is the study of space around a person with respect to others, was first defined by Hall [30]. It was found that proxemic zones are shaped by culture and psychophysical features ([31], [28], [29]). Schegloff [63] has shown that, in daily interactions, intentions are derived from the poses of the lower and upper parts of the human body, i.e. whether the involvement of the participant would be dominant or subordinate. Moreover, situational awareness is the ability to understand and perceive the environment around the person in order to plan and execute decisions and such relies on proxemics [24]. Thus, proxemics plays an important role in defining human-human interactions and relationships.

In the field of HRI, proxemics is vastly used for social navigation. For instance, [42], [19] used Hall's theory of proxemics to optimize social navigation of the robot while taking into account the human's safety and visibility. Additionally, proxemics can be used to initiate

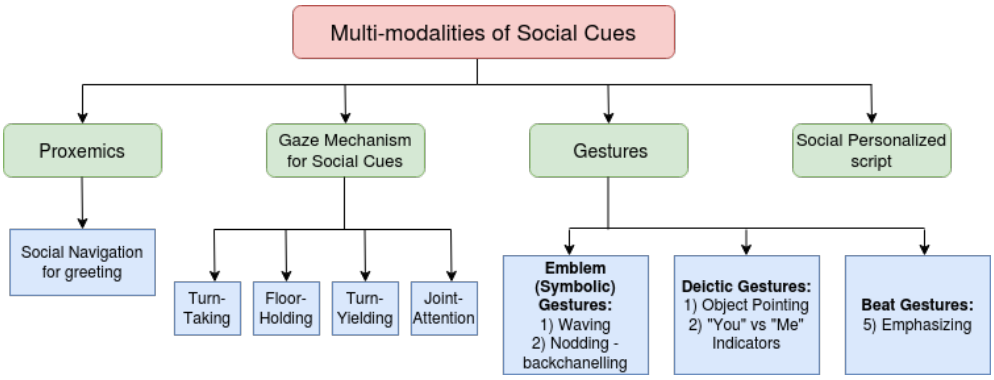


**Fig. 1** Summary of Multi-Modal Social Cues

interactions. Shi et al.[65] proposed a model based on proxemics and navigation to initiate a conversation with a human inspired by the study of human-human interactions. The robot with the implemented proxemics model was ranked higher in a subjective evaluation of appropriateness of initiation. An understanding of proxemics grants the robot a finer tool to perceive, predict, and manipulate the environment around the interaction and provides greater naturalness [65]. Furthermore, proxemics was used to estimate group formations around the robot allowing the robot to adapt its gaze based on the roles users are playing the group being formed, such as active participant, bystander, or overhearer [70]. In addition, a recent meta-analysis showed that for mobile robots, robot appearance or whether one is active or passive in the approach has no meaningful effects [46]. A pivotal question is whether proper use of this modality, which focuses on interpersonal space, would lead to stronger effects related to the outcomes of an interaction when compared to other modalities such as gaze and gesture.

### 2.4 Dialogue

In addition to the modalities of nonverbal behaviors, dialogue plays an influential role in forming impressions and manipulating social outcomes of the interaction. In human-human interactions, one study has shown that starting a conversation by asking people how they were feeling that day increases the likelihood of their compliance to a request for both charity donations and/or commercial purchases [21]. Moreover, in human-agent interactions, it was shown that having an agent start with a small request increased the chances of having the participant accept a bigger request shortly after [22]. Furthermore, verbal phrases influence social interactions with agents, e.g., separating emotional expressions targeted at an attitude versus at a person such as"your opinion" versus "you should" [77]. In HRI, dialogue similarly has an impact on the interactions including facilitating collaborations, managing errors, and personalizing conversations. For instance, it is used to exchange information and assist in human robot collaboration to achieve common goals [26]. Furthermore, social dialogue was shown to help robots recover from prior errors and gain future influence [48]. Additionally, service robots with personalized dialogues reinforced participants' rapport, cooperation, and engagement [44].

### 2.5 Multi-Modal Social Behaviors

Several studies have also combined two or more of these behaviors to assess the interaction between them. Most commonly, gaze behaviours are combined with gestures [32], proxemics ([70],[73], [25]) or verbal behaviours [15]. Works combining other modalities ([59], [56]), or comparing more than two or three modalities ([10], [38]) are rare. There is also a large body of work comparing 'social' and 'non-social' robots (or affective / emotional / personalized robots). However, the manipulations in these cases typically combine multiple modalities and evaluate overall system performance, as opposed to investigating the effect of specific modalities. As such, there is a clear need to develop a more comprehensive perspective on how different combinations of modalities (gaze, gesture, proxemics, and verbal content) contribute to overall perceptions of social intelligence during the course of an interaction.

## 3 Implemented System for Multi-Modal Behavior

This section introduces the implemented system to achieve multi-modal behavior on the Pepper robot (SoftBank Robotics). However, the system can also be implemented on other social robots. The source code for

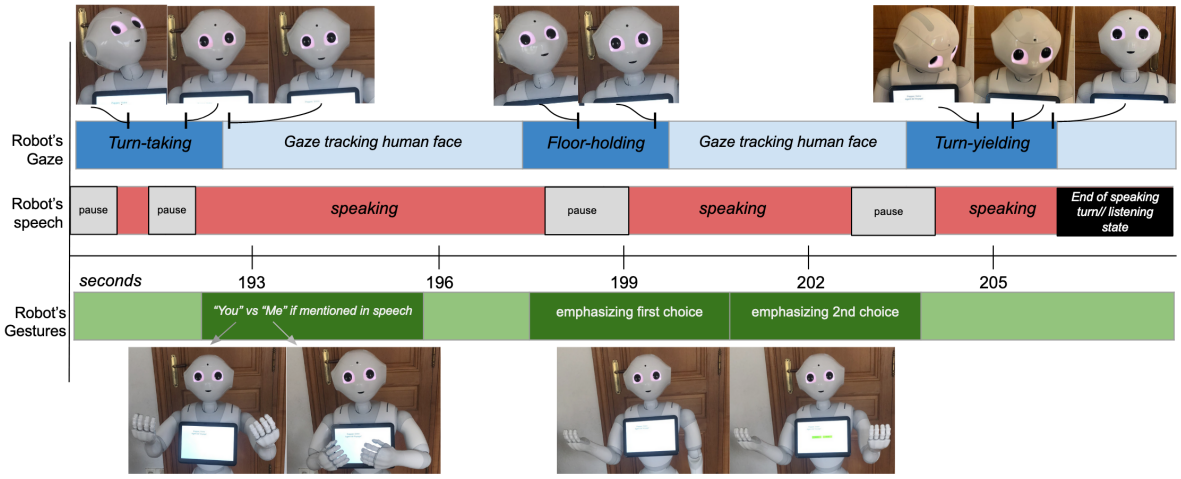


**Fig. 2** Sample of time-line including speech, gaze mechanisms (turn-taking, floor-holding, turn-yielding), and social gestures (deictic gestures: "You" vs "Me" if mentioned in speech, beat gesture: emphasizing the two choices user needs to select from)

the entire system has been made available online[2]. The overall scheme of the multi-modal social cues synthesized are shown in Figure 1. The system is composed of the following modalities: proxemics, gaze mechanism, gestures, and a social dialogue. A sample extract of the system implemented is shown in Figure 2.

### 3.1 Social Gaze Mechanisms

Since several humanoid robots lack expressive eyes that can be controlled, the presented social gaze aversions are achieved using head motion control. The social gaze mechanisms presented here were designed and implemented to fulfill the following functions: joint attention, turn-taking, floor-holding, and turn-yielding. When not performing these gaze aversions, the robot would be gaze tracking the human it is interacting with at all times. For this reason, information about the human can be extracted, notably the 3-dimensional frames of the human's face and the robot's gaze. This allows the robot to carry out all implemented gaze-averted head motions on the robot with respect to the frame of the human's face. Thus the design of each social gaze head movement was a combination of dynamics, magnitude, and duration, all of which are crucial for the social gaze motion to achieve its function naturally. A summary of the gaze mechanisms can be found in Figure 3.

Gaze aversion for turn-taking in human-human interaction as well as human-robot interaction holds a cognitive function; it gives the speaker more time to better plan and address their speech while also avoiding possible external distractions [7], [6]. For this reason, the turn-taking gaze behavior was given the relatively longer duration of 2.5 seconds. As for the floor management and turn-yielding functions, which take place during and at the of the speaking turn, they were assigned a shorter duration of 1.5 and 1.2 seconds, respectively. The longest duration was designed for the joint attention gaze, which has the duration of 3.8 seconds and the function of indicating and referring to an object of discussion. The angle rotations, duration length, and directions of the gaze aversions were selected and designed based on the gaze aversion system, with similar functionalities, implemented on the NAO robot (SoftBank Robotics) in [6], which are as well based on the findings of Kendon [37] for gaze aversions in human-human interaction. In turn, these variables were tuned to be suitable for the robot Pepper. Examples of how the duration of each gaze mechanism was synthesized is shown in Figure 2.

| Functions | Duration (seconds) | Direction | Magnitude |
|---|---|---|---|
| *Joint Attention* | 3.8 | Right | Head Yaw of -46° & Pitch of -4.6° |
| *Turn-taking* | 2.5 | Upward, left-sided | Rotation θ = 33.7° in the (y-z) plane of gaze frame |
| *Floor holding* | 1.2 | | |
| *Turn- Yielding* | 1.5 | Downward, Left-sided | x" = [(1−cosθ)]*x + y*sinθ - δ1<br>y" = -sinθ*x + (1- cosθ)*y<br>z" =δ2<br>θ = 12.318° , δ1= 1.2, and δ2=−0.6<br>*x, y,& z are the coordinates of human head frame* |

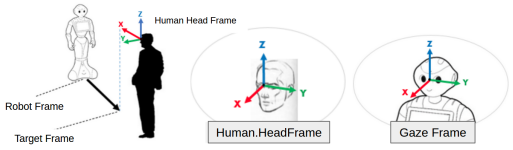


**Fig. 3** Summary of gaze mechanisms

[2] Multi-modal Social Cues System Implementation GitHub Repository `https://github.com/KarenTatarian/multimodal_socialcues`

The goal is to generate head motions, which serve as social gaze cues, and to realize a natural and subtle appearance while still fulfilling their purpose. The directions of the head rotations performed were chosen based on human-human as well as HRI studies done to better understand gaze aversions during interactions ([6] [36] [53]). For the turn-taking and floor-holding cues, which signify that the robot is in a cognitive phase and holding the speaking role, a vertical head movement was implemented. The gaze frame of the robot is the frame it produces as it is tracking the human's face as seen in Figure 3, thus the gaze frame is aligned with the human's face frame. At each instance in which the turn-taking and floor-holding gaze mechanisms were called, the gaze frame performs a rotation of 33.7 degrees in the (y-z) plane with respect to the frame on the human's face resulting in an upward slightly left-sided rotation. As for the turn-yielding gaze, in order to achieve a smoother and more subtle gaze to indicate the shift in speaking roles and after manually fine tuning, Equation 1 represents the translations implemented for the $(x, y, z)$ of the robot's gaze frame respectively

$$\begin{aligned} x" &= [(1 - \cos\theta)]x + \sin\theta y - \delta_1 \\ y" &= -\sin\theta x + (1 - \cos\theta)y \\ z" &= \delta_2 \end{aligned} \tag{1}$$

where $\theta = 12.318°$, $\delta_1 = 1.2$, and $\delta_2 = -0.6$ and $x, y, \& z$ are the coordinates of the frame of the human's face.This resulted in a downward left-sided rotation as seen in Figure 4, which illustrates an example of the execution. Finally, the head rotation for joint attention was not intended to be subtle but rather draw the attention of the user towards the object it is referring to. The head rotation was done with a head yaw of $-46°$ and a head pitch of $-4.65°$ with respect to the object it is referring to and with a Bézier curve velocity profile for smoothness in the animation. This resulted in an animation that starts slow, speeds up during the main shift, and finally slows down before it ends. For instance, for the gaze mechanisms implemented as head-motions, the head of the robot would start rotating slowly then rapidly towards target frame then slowly returns back to its initial gaze frame as shown in the example of Figure 4.

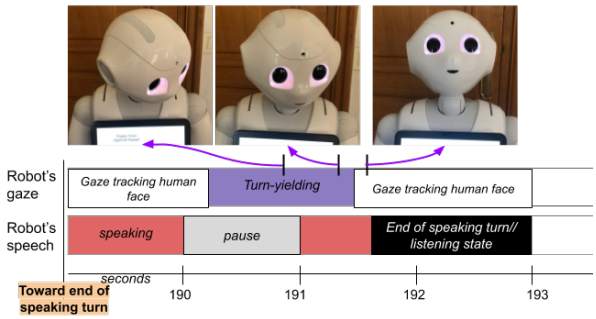


**Fig. 4** Turn Yield Gaze Mechanism

## 3.2 Gestures

Three of the five categories of gestures classified in [49],[41], are implemented in this multi-modal system through five gestural designs. First, in the emblem category, two gestures were implemented: the *wave gestures* for greeting users and closing interaction and *nodding* as back-channeling. When the robot enters a listening state, it nods in response to additional information the user provides while speaking. Whenever a new sentence is detected by the robot, it executes a back-channeling nod animation and waits for a next sentence to be heard. The back-channeling nod animation consists of three consecutive nods over a span of 2.5 seconds. Second, in the deictic category, two types were represented: *object pointing gesture*, which accompanied the joint attention gaze and performed as a right arm extension, and *self/other references*, which are the "you" vs "me" indicators, as portrayed in Figure 2. In the latter, the "you" gesture was designed as both of the arms open and extended from the robot's body and pointing upwards towards the human with the open palms. The "you" gesture accompanied questions,in which the robot infers about the user. On the other hand, the self-pointing or the "me" gesture has the robot's arms closed and directed towards the body and it is used by the robot to introduce itself, its tablet modality, and state an opinion. Furthermore, beat gestures were used in the question series to emphasize the options. In order to generate smooth behavioral social cues, Bézier curves were used to implement the velocity changes of all gestures.

## 3.3 Proxemics for Approaching Human

According to the findings in ([67],[42]), the most optimal and socially acceptable path to navigate towards a static standing human is in a straight line in the human's sight zone while maintaining a minimum distance of about 0.5 meters (or 1.5 meters for sitting humans). This is due to safety and visibility constraints to minimize human discomfort. In addition, approaching the human for initiating an interaction was ranked as most appropriate when the robot navigated in the human's front zone (120° cone-shaped area in front of a person's head) or his/her gaze zone (30°) [65]. For this system, once the human has been detected at the beginning of the interaction,

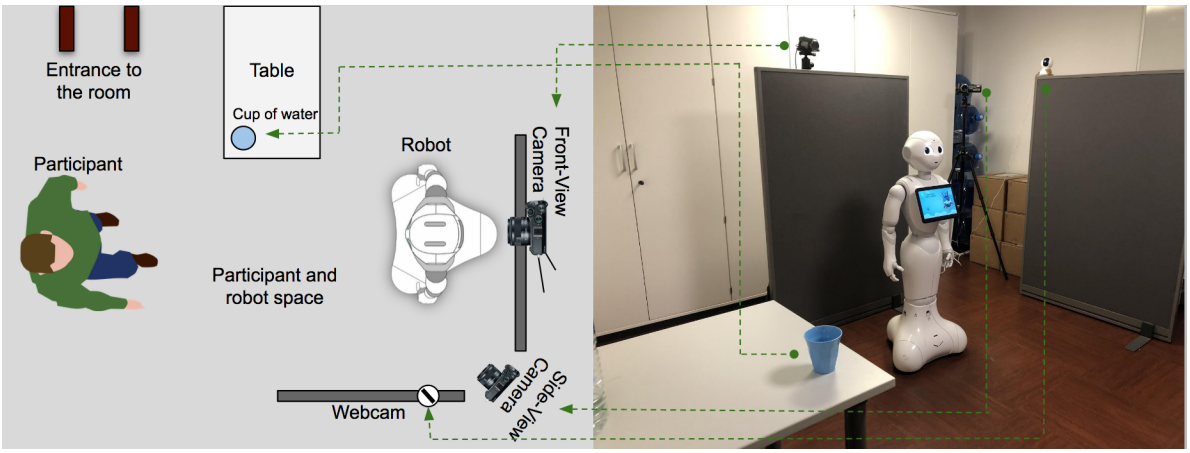


**Fig. 5** Schematics of the experimental room set-up with the robot during the travel agent scenario

the robot would first greet the human by speech and a wave gesture before beginning navigation towards the human. The default speed of the robot Pepper is 0.35 m/s. However to avoid a recoil movement by the human seen in some user experience testing, the speed for greeting navigation was slowed down to 0.25 m/s. The robot would then navigate to establish a distance of 0.85 meters between itself and the user. The robot would navigate while maintaining gaze directed towards the participant. The distance of 0.85 meters was chosen for four main reasons. First, this distance allows the robot to continuously track the human's face regardless of their height. Second, the human at this distance is able to clearly see the different gaze and gesture behaviors generated by the robot. Third, this distance eases the access and view of the tablet on the robot. Fourth, 0.85 meters is still within Hall's defined personal distance, in which friendly interactions take place [29]. Once the desired distance has been reached, whether by navigation of the robot or chosen distance by human, the start button pops up on the tablet to continue the rest of the interaction.

## 4 Design Method and Evaluation

In order to directly compare the effects the different modalities of the robot's behavior have on the user's behavior and attitude, the chosen scenario for the interaction was planning for a hypothetical holiday (Fig. 5). The robot acted as a travel agent helping the user plan their next vacation and it exhibited one of five behavioral conditions:

- Multi-modal Interaction, which is all modalities including social dialogue, *(Social Gaze + Gestures + Proxemics + Social Dialogue)*
- Minus Proxemics, which is all the modalities excluding proxemics, *(Social Gaze + Gestures + Social Dialogue)*
- Minus Social Dialogue, which is all the modalities excluding social dialogue*(Social Gaze + Gestures + Proxemics )*
- Minus Gestures, which is all the modalities excluding social gestures, *(Social Gaze + Proxemics + Social Dialogue)*
- Minus social gaze, which is all the modalities excluding social gaze, *(Gestures + Proxemics + Social Dialogue)*

The flow of interaction went as seen in Figure 6: first in the introductory phase, the robot greeted the user (highlighting the gesture modality), approached the human until the desired distance is established, engaged the user in a short social dialogue and then offered the participant some drinking water, using joint attention gaze and pointing gesture. Second was the travel planning phase, where the robot started asking a series of questions about the travel and vacation preferences, followed by a self-disclosure segment, where an open ended question was asked to know more about the user and the robot entered a listening state and demonstrated backchanneling. The third and final part was the closing phase, where the robot suggested two options based on answers previously provided by participants and recommended its personal preference between the two. Once the user made a final decision on a travel destination, the robot concluded the interaction and waved goodbye.

### 4.1 Social Dialogue

The dialogue throughout all conditions was adaptive: the robot's answers depended on what the user's previous choices were. In addition, since it was a travel planning

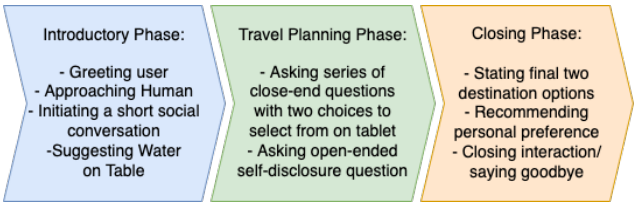


**Fig. 6** Summary of flow of interaction

| **Questions** *Used, French* | **Questions** *English* |
|---|---|
| Pepper est gentil | Pepper is friendly |
| Pepper est chaleureux | Pepper is warm |
| Pepper est aimable | Pepper is likeable |
| Pepper est accessible | Pepper is approachable |
| Je demanderais volontiers des conseils à Pepper | I would ask Pepper for advice |
| J'aimerais avoir Pepper comme collègue | I would like Pepper as a colleague |
| J'aimerais avoir Pepper comme colocataire | I would like Pepper as a housemate |
| J'aimerais que Pepper et moi soyons amis | I would like to be friends with Pepper |
| Pepper et moi sommes similaires | Pepper is similar to me |

**Table 2** Likeability Scale ranked from 1 to 7

scenario, all replies from the robot were consistent with the decisions the user made. For instance, if the person selected that they prefer to travel by train then the robot would suggest a destination that can be reached from Paris by train (e.g., Amsterdam) and similarly for the other descriptions, e.g., city or beach, solo trip or with friends and loved ones, culture or activities etc. However, the social dialogue modality differs in the social and friendly openings and replies as summarized in Table 1.

| | Conditions | |
|---|---|---|
| | Multi-modal Interaction, Minus Proxemics, Minus Gestures, and Minus Social Gaze | Minus Social Dialogue |
| **Openings/Closing** | Social small talk<br>*ex: "How are you today?" ... "great"*<br>*ex: "I am so happy to meet you!"* | Formal talk<br>*ex: "Is it the first time you come here?"*<br>*ex: "You are the fourth person today whom I will help plan their vacation"* |
| **Replies** | Personal preference replies<br>*ex: "Excellent choice, I also like this location!"*<br>*ex: "I also find it awesome to travel by train because it's much more comfortable!"* | Non-personal general replies<br>*ex: "Many people like this location"*<br>*ex: "Traveling by train is more comfortable"* |

**Table 1** Social dialogue designs per condition

## 4.2 Design & Materials

An independent groups design was used, with the independent variable being multi-modal behavior with 5 levels: first: multi-modal interaction referring to all implemented modalities (proxemics, social gaze, gestures, and social dialogue), second: minus proxemics - referring to all implemented modalities except for proxemics, third: minus social dialogue - referring to all implemented modalities except for social dialogue, fourth: minus gestures - referring to all implemented modalities except for gestures, and fifth: minus social gaze - referring to all implemented modalities except for social gaze mechanisms.

The dependent variables were extracted using recorded logs from the robot application, data extracted from the recorded videos, and self-report questionnaires. First, the logs provided from the robot application include information on the position of the human relative to the robot extracted every 5 seconds as well as the angular facial frame information, which were extracted before the execution of every social gaze aversion. In addition, the time it took the user to press the buttons on the tablet and to take decisions as well as the decisions made were recorded.

Second, the videos recorded were used to annotate and obtain the verbal and nonverbal responses and behaviors of the user throughout the interaction, including if the user accepted the water offer, back-channels performed, verbal responses, total speaking time, amount of information shared, number of audio/voice recognition errors that may have occurred, and gestures performed. It is critical to note that the back-channels of the users in this set-up refer to both non-lexical back-channels, such as "uhh", "yeah", "mmm", .. etc., phrasal back-channels, such as "wow", "great",.. etc., and gestural back-channels, e.g., nodding. However, facial expressions were not considered.

Third, a self-report questionnaire was used to evaluate perceived agency, social trust, competency trust, liking, rapport, acceptance, social presence, and social information processing. This paper will only discuss the results of the *Likeability scale* of the questionnaire. The subjective measurement for robot's Likeability or *liking* was comprised of 9 items. The participants were asked to rank how well they agree with the statements in Table 2 on a scale of 1 to 7, with 1 agree with the least and 7 agree with the most. These items were then averaged to form a scale with good reliability $\alpha = 0.88$. The scale was normally distributed, $W = 0.99, p = 0.35$.

## 4.3 Participants & Procedure

115 participants were recruited for the experiment, of which 10 had to be excluded due to technical difficulties with the robot. Thus the data of 105 participants were used for analysis ($mean\ age = 22.8$, $SD = 3.17$) with ($N = 21$) participants per condition. All participants were recruited by the INSEAD-Sorbonne University

Behavioural Lab under ethics approval by INSEAD Institutional Review Board. All participants were native French speakers and signed a consent form to participate. Separate consent was obtained for the use of video data. The entire experiment took about 20 minutes to complete, including filling the questionnaires at the end. As a compensation for their time each participant received 6 euros.
Participants were randomly assigned to one of the five conditions (21 per condition). In addition, the participants were assigned randomly across the hours of the day to make sure half of the users in each condition interacted in the morning before lunch time while the other half in the afternoon after lunch.
Upon arrival, the participants filled out the consent forms. Then, one of the experimenters introduced the experiment by explaining that they will interact with a robot that will help them plan their vacation as if it were their real holiday. They were informed the interaction would last 5-7 minutes and then they would have to fill a questionnaire, which included the liking scale as well as scales not relevant to the current study. They were also advised to speak loudly and articulate clearly in order to avoid any audio or voice recognition problems. Participants were then led to a room with the robot, as seen in Figure 5, and were asked to place their belongings on the side and stand wherever they wished. Shortly after the experimenter leaves the room, the interaction began. Besides the front and side cameras set-up in the room, there was a webcam streaming the interaction live. The robot as shown previously is completely autonomous, the videos and webcam are for recording and monitoring. Once the interaction was over, the participants were asked to fill the questionnaire in a different room and then they were debriefed on the study and were given their participation compensation.

### 4.4 Hypotheses

To evaluate the modalities in a multi-modal interaction and their effects on perceived social intelligence, the following hypotheses were formulated. First, the proxemics implemented respected the personal distance established by [29] and visibility and safety [42] while still initiating the interaction by approaching the user within his/her gaze zone [65]. Following this, H1 was suggested:

- H1: Social distances established by the robot would be maintained throughout the interaction in all conditions except *Minus Proxemics*.

Second, social gaze aversion was shown to play a major role in *intimacy* regulation during human-human interactions to elevate the comfort of speakers ([36], [8]). In addition, gaze aversion for turn-taking functioned as a social cue to hand the conversational floor to the user and thus making him/her the speaker. Furthermore, gaze aversion is practiced by humans specially when listening in order to minimize the negative perception attributed to staring and to promote the comfort of the speaker ([1], [17] ). While Andrist et al. [6] did not find that a social robot with proper timings for gaze aversions increased self-disclosure and comfort in humans more than a social robot with badly timed gaze aversions, we hypothesise that gaze mechanisms supported by multimodal behaviors would elicit more self-disclosure from the participants, such that:

- H2: Time participants spend speaking in the self-disclosure segment would be the shortest in the *Minus Gaze* condition relative to the other conditions.

Third, gesture and joint attention through gaze have shown to be modalities used to communicate and point at an object of reference in an interaction as well as asking to grab the object referred to ([12], [5] [68]). With H3 formulated as:

- H3: Water suggestions are more likely to be taken when participants interact with a robot performing *social gaze mechanisms* and *gestures*.

Fourth, gestural alignment was proposed to measure the extent to which an interaction shapes the behavior of the user and the smoothness of the interaction ([14], [16]). As such, the following hypotheses were formulated looking into gestural alignment at the greeting and termination phases of the interaction in order to understand the possible change in gestural alignment behavior of the users. The use of back-channels, which include nodding and verbal content, throughout the interaction were also analysed as part of gestural alignment:

- H4a: Gestural alignment in the greeting and termination phases would be least present in the *Minus Gesture* condition.
- H4b: *the complete multi-modal behaviour* condition would have the most participants who at the beginning did not greet the robot but at the end did close the interaction with the robot whether verbally or non-verbally.
- H4c: Back-channeling throughout the interaction would be least performed by participants in the *Minus Gesture* condition.

Fifth, all the modalities combined make up multi-modal social cues designed to facilitate a more natural and friendly interaction. It was hypothesized that that would additionally have an effect on the subjective attitude of the users.

– H5: The *condition with all modalities and social dialogue* would score higher on the likeability scale questionnaire.

## 5 Results

### 5.1 Distances from Pepper

First, looking into the distances established throughout the interaction, one would expect no difference in the initial distance, as no interaction has yet occurred, but rather a difference after the navigation of the robot. We therefore conducted Kruskal-Wallis H-tests for the average distance maintained from the robot at the beginning of the interaction (prior to the social navigation phase), shown in Figure 7, as well as the average distance maintained after the social navigation phase until the end of the interaction, shown in Figure 8. The former model yielded no significant difference for condition, however, the latter revealed a significant effect of condition after the social navigation phase, see Table 3.

**Table 3** Kruskal-Wallis Tests for the Effect of Condition on Initial and Maintained Distances

| Outcome | $\chi^2$ | $df$ | $p$ | $\epsilon^2$ |
|---|---|---|---|---|
| Initial Distance | 2.40 | 4 | .663 | 0.02 |
| Distance Maintained | 23.92 | 4 | <.001 | 0.23 |

Follow up pairwise comparisons with Dwass-Steel-Critchlow-Flinger correction revealed participants in the minus proxemics condition stood significantly further away from the robot throughout the interaction than in all other conditions (all $W's \geq 4.68$, all $p's < 0.008$). No other differences between conditions were significant. This denotes that participants maintained the close social distance due to the proxemics established by the robot, thus validating **H1**.

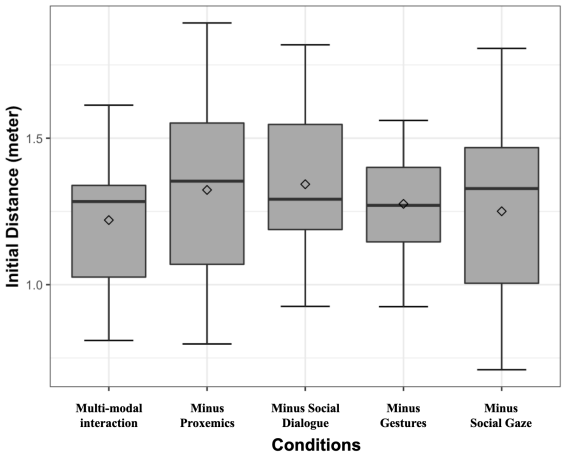


**Fig. 7** Initial Distance (meters) by participants per condition at the beginning of the interaction with the robot before proxemics

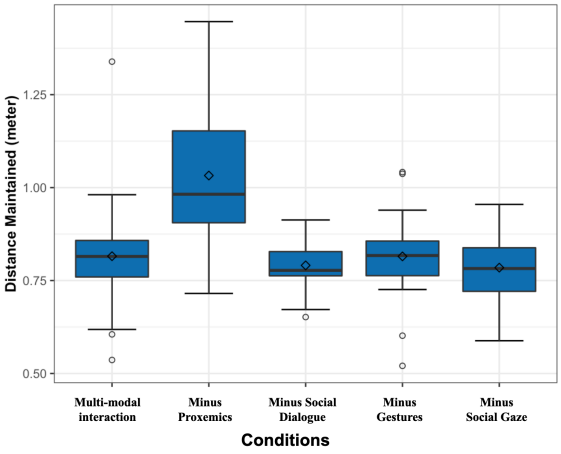


**Fig. 8** Maintained Distance (meters) by participants per condition throughout the interaction with the robot after proxemics.

### 5.2 Self-Disclosure

Second, we examined the effect of modalities on self-disclosure by looking into how comfortable a human was in sharing information about themselves when asked by the robot. We again conducted a Kruskal-Wallis H-test to evaluate the total speaking time (in seconds) the user spent answering the robot's open-ended question to talk about themselves, shown in Figure 9. The average total speaking time was $M = 11.55, (SD = 6.05)$ seconds. However, there was no significant effect of modality, $\chi^2(4) = 4.09, p = .394, \epsilon^2 = 0.04$. As such, **H2** was not supported. The number of pieces of information given, which is the number of new facts or opinions revealed and provided by the user about him/herself, was annotated and measured for each condition. The following is an example of how the data was annotated: if a participant after the open-ended self-disclosure question answered "I like hanging out with my friends...I like watching movies", then this was annotated as two pieces of information since two facts and/or opinions were revealed about the participant. The average number of pieces of information given was $M = 3.429, (SD = 1.6)$, but there was again no significant effect of condition, $\chi^2(4) = 1.76, p = .780, \epsilon^2 = 0.02$.

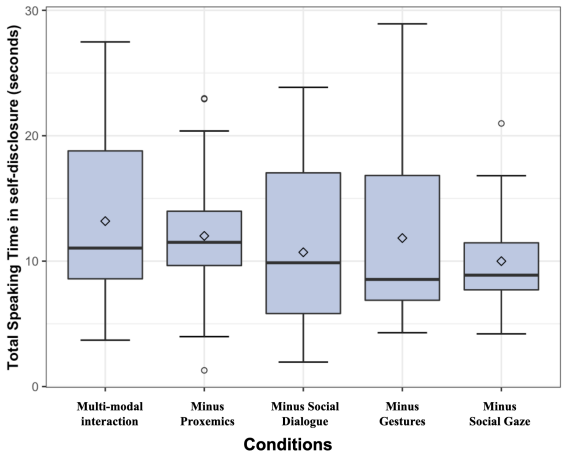


**Fig. 9** Total speaking time (seconds) of participants per condition during self-disclosure open-ended question

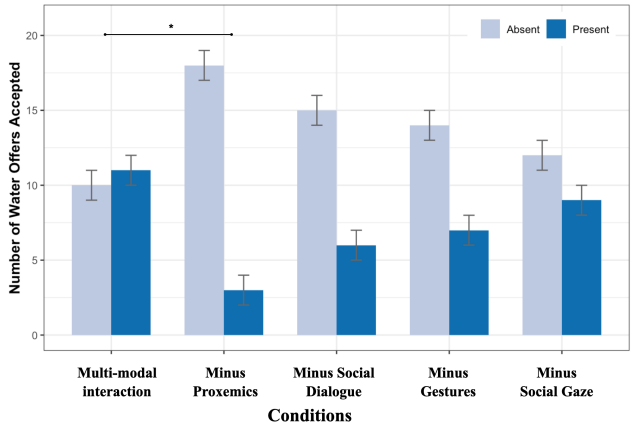


**Fig. 10** Number of times the water offer was accepted by participants per condition. *Present* refers to appearance of the participant's behavior of accepting the water offered by the robot by grabbing the cup and/or drinking the water, while *Absent* refers to lack of this behavior and as such not accepting the water offered by the robot.

## 5.3 Accepting Water Offered

The number of participants who accepted the water offered in each condition is shown in Figure 10. It is important to note that *absent* in this figure and those that follow refer to the number of times the observed behavior was absent in each condition. For instance, in Figure 10, the grey or *absent* plots indicate the number times the water suggestion was *not* taken (and as as such absent). We ran a binomial logistic regression with condition as the predictor variable, and accepting the water offered as an outcome variable. The overall test of condition was marginally significant, $\chi^2(4) = 8.18, p = .09$, McFadden's Pseudo $R^2 = 0.06$. Follow up pairwise comparisons with Tukey's correction revealed a marginally significant difference between the complete multimodal condition and the minus proxemics condition, $\beta = 1.89, p = .095$. The odds of accepting the water offered in the minus proxemics condition were 0.15 [0.03, 0.62] times less than in the complete multi-modal condition.

## 5.4 Social Behavior

We again constructed binomial logistic regression models assessing the effect of condition on opening and closing waves. For both opening and closing waves, the logistic regression model was significant for condition, $\chi^2(4) = 19.20, p < .001$, McFadden's Pseudo $R^2 = 0.13$ and $\chi^2(4) = 35.54, p < .001$, McFadden's Pseudo $R^2 = 0.25$, respectively. We conducted follow up pairwise comparisons with Tukey's correction. As the closing wave model exhibited complete separation (i.e., no participants in the minus gesture condition waved goodbye), we further applied Firth's bias reduction method for this model. For both opening and closing waves, participants were significantly less likely to wave in the minus gesture condition than in the complete multi-modal condition and the minus Social Dialogue condition (see Table 4). No other comparisons were significant. See Figures 11 and 12.

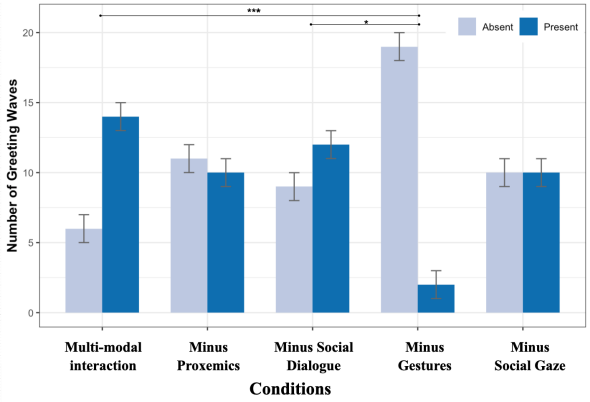


**Fig. 11** Number of *greeting* waves performed by participants at the beginning of the interaction (with 95% CI errors). *Present* refers to appearance of the participant's behavior of waving to the robot at the beginning of the interaction, while *Absent* refers to lack of this behavior, e.g., the participants not waving to the robot.[*] significant at the $p < .05$ level. [***] significant at the $p < .001$ level

Beyond waving gestures in greeting and closing the interaction, Table 5 shows the presence and absence of all greeting and/or closing turns made by users, whether verbally or non-verbally. We classified participants behaviour as either consistent-social (both greeting and closing turn). consistent-nonsocial (neither greeting nor closing), inconsistent-social (no greeting turn, but a closing turn) or inconsistent-nonsocial

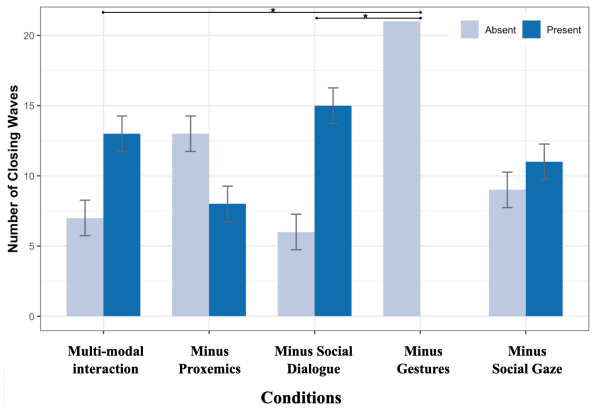


**Fig. 12** Number of *closing* waves performed by participants at the end of the interaction(with 95% CI errors). *Present* refers to appearance of the participant's behavior of waving to the robot at the end of the interaction, while *Absent* refers to lack of this behavior, e.g., the participants not waving to the robot.[*] significant at the $p < .05$ level

**Table 4** Comparison of the likelihood of producing opening and closing waves in each condition

| Contrast | Log Odds (SE) | z | p-value |
|---|---|---|---|
| **Opening Wave** | | | |
| Minus Gesture vs Multi-Modal Interaction | -3.10 (0.89) | -3.49 | .005** |
| Minus Gesture vs Minus Gaze | -2.25 (0.87) | -2.60 | .071 |
| Minus Gesture vs Minus Social Dialogue | -2.54 (0.86) | -2.94 | .027* |
| Minus Gesture vs Minus Proxemics | -2.16 (0.86) | -2.50 | .090 |
| Multi-Modal Interaction vs Minus Gaze | 0.85 (0.66) | 1.28 | .704 |
| Multi-Modal Interaction vs Minus Social Dialogue | 0.56 (0.66) | 0.85 | .915 |
| Multi-Modal Interaction vs Minus Proxemics | 0.94 (0.66) | 1.44 | .602 |
| Minus Gaze vs Minus Social Dialogue | -0.29 (0.63) | -0.46 | .991 |
| Minus Gaze vs Minus Proxemics | 0.10 (0.63) | 0.15 | .999 |
| Minus Social Dialogue vs Minus Proxemics | 0.38 (0.62) | 0.62 | .972 |
| **Closing Wave †** | | | |
| Minus Gesture vs Multi-Modal Interaction | -4.35 (1.54) | -2.83 | .038* |
| Minus Gesture vs Minus Gaze | -3.95 (1.53) | -2.58 | .074 |
| Minus Gesture vs Minus Social Dialogue | -4.63 (1.54) | -3.01 | .022* |
| Minus Gesture vs Minus Proxemics | -3.30 (1.53) | -2.15 | .197 |
| Multi-Modal Interaction vs Minus Gaze | 0.40 (0.65) | 0.61 | .973 |
| Multi-Modal Interaction vs Minus Social Dialogue | -0.28 (0.67) | -0.42 | .994 |
| Multi-Modal Interaction vs Minus Proxemics | 1.05 (0.65) | 1.62 | .482 |
| Minus Gaze vs Minus Social Dialogue | -0.68 (0.66) | -1.03 | .840 |
| Minus Gaze vs Minus Proxemics | 0.65 (0.64) | 1.03 | .842 |
| Minus Social Dialogue vs Minus Proxemics | 1.33 (0.66) | 2.03 | .251 |

† With Firth's bias reduction method
* significant at the $p < .05$ level
** significant at the $p < .01$ level

(greeting turn, but no closing turn). However, a chi-square test comparing participants consistency in social behaviour did not reveal any differences between conditions, $\chi^2(16) = 13.11, p = .664$, Kramer's V $= 0.194$ $[0.00, 0.33]$.

In addition, to assess the effect of modalities on behavioral alignment, we analysed the back-channeling performed by the participants in each condition. A binomial logistic regression was conducted to analyze the effect of the modalities on the number of participants who performed back-channeling, as shown in Figure 13. The regression model was statistically significant with $\chi^2(4) = 12.90$, $p = .012$, McFadden's Pseudo $R^2 = 0.09$. Evaluation of the log odds with Tukey's correction revealed participants in the minus gaze condition were less likely to produce back-channels than in the complete multi-modal condition, see Table 6.

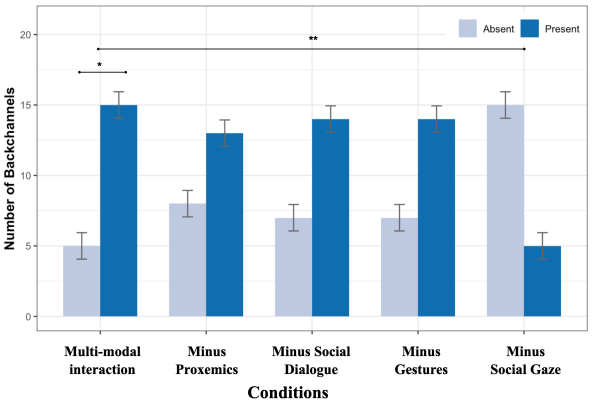


**Fig. 13** Number of back-channels performed while interacting with the robot in each condition (with 95% CI errors). *Present* refers to appearance of the back-channels performed by the participants, while *Absent* refers to lack of back-channels detected during the interaction.[*] significant at the $p < .05$ level. [**] significant at the $p < .01$ level

## 5.5 Liking

A Kruskal-Wallis H-test for participants subjective evaluations of their liking of the robot revealed no differences between any of the conditions, shown in Figure 14, $\chi^2(4) = 3.84, p = .428, \epsilon^2 = 0.04$; failing to support **H5**. Further exploratory investigation was done to look into behavioral outcomes that might represent liking. First, Figure 15 shows the number of addressee terms used to address the robot in each condition. In the french language, the pronoun "tu" (referring to "you") is used in informal and/or friendly contexts, whereas, the pronoun "vous" is used for formal and/or acquaintance contexts. See Table 7 for the frequency of each mode of address used by participants. A chi square test comparing participants mode of address towards Pepper in each condition was significant $\chi^2(12) = 23.01, p = .028$, Kramer's V $= 0.27, [0.00, 0.32]$. Follow up tests with FDR correction, however, did not reveal specific differences between conditions (Table 8). As this analysis was exploratory, we then relaxed the need for correction with multiple comparisons. Without correction, there was a significant difference between the minus Social Dialogue and minus proxemics conditions.

Second, Figure 16 shows in each condition the number of participants who made utterances while using the tablet (despite it being clear that speech was not needed

| **Status of Greeting/Closing:** | **Multi-modal Interaction** ($n_1$) | **Minus Proxemics** ($n_2$) | **Minus Social Dialogue** ($n_3$) | **Minus Gestures** ($n_4$) | **Minus Gaze** ($n_5$) | **Total (N)** |
|---|---|---|---|---|---|---|
| **Greeted** robot & **Closed** Interaction | 10 | 15 | 16 | 13 | 14 | 68 |
| **Greeted** robot but **Did not Close** Interaction | 4 | 2 | 1 | 1 | 2 | 10 |
| **Did not Greet** robot but **Closed** Interaction | **5** | **3** | **1** | **5** | **2** | 16 |
| **Did not Greet** robot & **Did not Close** Interaction | 1 | 1 | 3 | 2 | 1 | 8 |

**Table 5** Number of greeting and/or closing turns, which may be verbal or non-verbal, present and absent in each modality in beginning and end of the interaction. The rows represent the statuses in each condition as follows: 1) participants who greeted the robot and also closed the interaction 2) those who greeted the robot but did not close the interaction 3) those who did not greet the robot but closed the interaction 4) those who did not greet the robot nor did they close the interaction

**Table 6** Comparison of the likelihood of producing back-channels in each condition

| **Contrast** | **Log Odds (SE)** | ***z*** | ***p-value*** |
|---|---|---|---|
| Minus Gesture vs Multi-Modal Interaction | -0.41 (0.69) | -0.59 | .977 |
| Minus Gesture vs Minus Gaze | 1.79 (0.69) | 2.58 | .073 |
| Minus Gesture vs Minus Social Dialogue | 0.00 (0.66) | 0.00 | 1.00 |
| Minus Gesture vs Minus Proxemics | 0.21 (0.65) | 0.32 | .998 |
| Multi-Modal Interaction vs Minus Gaze | 2.20 (0.73) | 3.01 | .022* |
| Multi-Modal Interaction vs Minus Social Dialogue | 0.41 (0.69) | 0.59 | 0.98 |
| Multi-Modal Interaction vs Minus Proxemics | 0.61 (0.69) | 0.80 | .899 |
| Minus Gaze vs Minus Social Dialogue | -1.79 (0.69) | -2.58 | .073 |
| Minus Gaze vs Minus Proxemics | -1.58 (0.69) | -2.31 | .140 |
| Minus Social Dialogue vs Minus Proxemics | 0.21 (0.65) | 0.32 | .998 |

* significant at the $p < .05$ level

**Table 7** Frequency of modes of address used by participants towards Pepper in each condition

| **Condition** | **Mode of Address** | | | |
|---|---|---|---|---|
| | **Tu** | **Vous** | **Pepper** | **Nothing** |
| Multi-Modal Interaction | 1 | 1 | 0 | 19 |
| Minus Gaze | 1 | 3 | 0 | 17 |
| Minus Gesture | 2 | 3 | 0 | 16 |
| Minus Social Dialogue | 0 | 1 | 3 | 17 |
| Minus Proxemics | 5 | 1 | 0 | 15 |
| **Totals** | 9 | 9 | 3 | 83 |

to carry out choice selection on the tablet of the robot). For instance, listening mode and speech recognition in the robot was activated when the robot asked open ended questions; whereas for making choices regarding the planning of the destination in this scenario, the interface to select the preference was by using the tablet (as was instructed by the robot in the beginning). There was no significant difference in how likely participants were to talk to Pepper in addition to using the tablet, $\chi^2 = 5.18, p = .270$, McFadden's Pseudo $R^2 = 0.04$.

**Table 8** Comparison between modes of address used by participants towards Pepper in each condition

| **Contrast** | **Raw *p*-value** | **FDR-corrected *p*-value** |
|---|---|---|
| Multi-Modal Interaction vs Minus Gaze | .798 | .886 |
| Multi-Modal Interaction vs Minus Gesture | .593 | .740 |
| Multi-Modal Interaction vs Minus Social Dialogue | .232 | .386 |
| Multi-Modal Interaction vs Minus Proxemics | .229 | .386 |
| Minus Gaze vs Minus Gesture | 1.00 | 1.00 |
| Minus Gaze vs Minus Social Dialogue | .223 | .386 |
| Minus Gaze vs Minus Proxemics | .207 | .386 |
| Minus Gesture vs Minus Social Dialogue | .122 | .386 |
| Minus Gesture vs Minus Proxemics | .347 | .496 |
| Minus Social Dialogue vs Minus Proxemics | .022* | .222 |

* significant at the $p < .05$ level

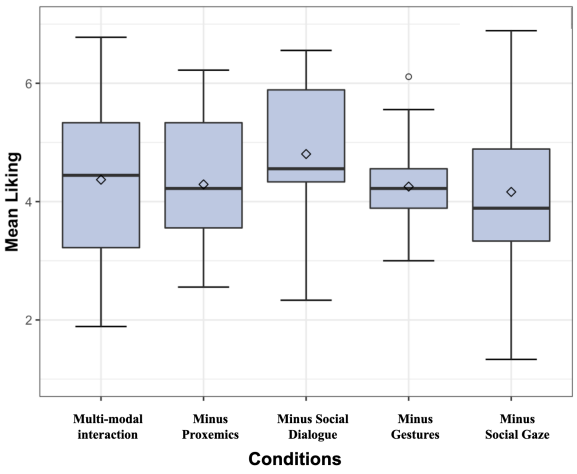


**Fig. 14** Mean Liking ranked by participants for the robot in each condition

## 5.6 Voice Recognition Errors

Although we attempted to limit the amount of autonomous voice recognition, the introductory phase included some reciprocal interaction between Pepper and the participant (e.g., asking "how are you"). Thus, there was still some potential for voice recognition errors to

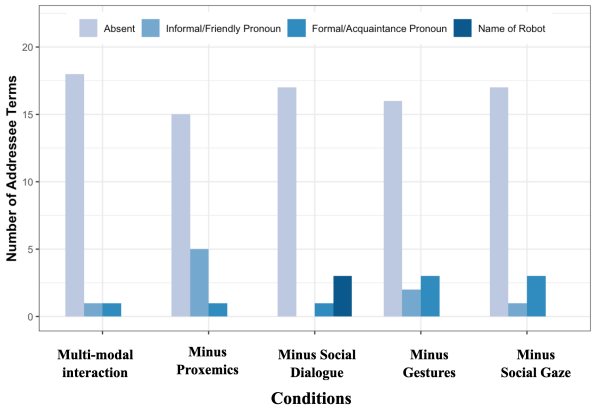


**Fig. 15** Number of addressee terms used to address the robot in each condition. The addressee terms are placed in four categories: absent, no term was used to directly address the robot, informal/friendly pronoun, formal/acquaintance pronoun, and name of the robot

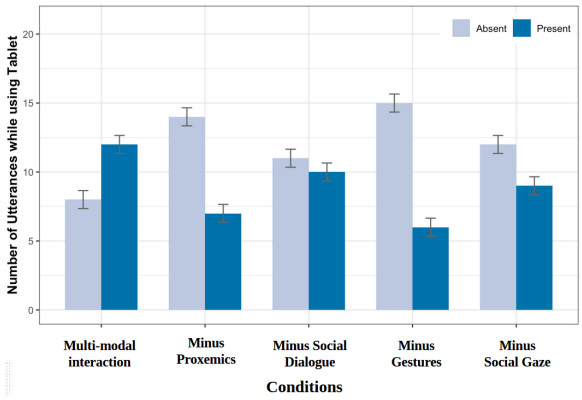


**Fig. 16** Number of Utterances while using the tablet of the robot in each condition.*Present* refers to detection of utterances/speech while tablet was used by the participants, while *Absent* refers to lack of detection.

occur. Although not initially part of the experimental design, based on observations of instances where voice recognition errors occurred we decided to to explore if these (naturally occurring) errors had any effect on participants behaviour.

A one-way Kruskal Wallis H-test performed on the number of voice recognition errors occurring per condition was non-significant, $\chi^2(4) = 2.04$, $p = .727$, $\eta^2 = 0.02$, indicating there was no difference in the number of voice recognition errors occurring between conditions.

Significant negative correlations were identified between the number of voice recognition errors, the total time the participant spent talking during the self-disclosure phase $\rho = -0.20$, $p < .05$ and the number of pieces of information they disclosed $\rho = -0.28$, $p < .01$.

## 6 Discussion

First, *H1* was supported, showing the influence of proxemics on the distances maintained by the users throughout the interaction. Prior to the navigation of the robot, participants chose to stand far from the robot at a ($M = 1.28, SD = 0.26$) meters distance, with no significant difference between conditions. The initial distance chosen was within the social distance defined by Hall [29] and did not give accessibility to the robot's tablet. In the conditions where the robot navigated to establish the personal distance of 0.85 meters, distances maintained during the rest of the interaction until the end were much closer. Conversely, in the *minus proxemics* condition, participants kept a further distance from the robot. Proxemics once again played an influential role on the behavioral outcomes of the interaction and was the main reason users kept a close distance to the robot.

Second, *H2* was not supported, *social gaze* mechanisms did not elicit an increase in self-disclosure speaking time nor the amount of information the user revealed about themselves. Further investigation was held to interpret what might have affected the self-disclosure speaking turn of the users. It was found that *voice recognition errors* significantly predicted total speaking time of participants and the amount of information shared. While gaze aversions and their respective functions play a guiding role in intimacy regulation and comfort in self-disclosure in human-human interactions, findings in this study seem to show that for human-robot interactions technological voice recognition errors precede gaze aversions in governing behavioral outcomes for such contexts. This shows that getting the robot technologically ready may have a great impact on how naturally a user answers an open ended question about themselves rather than how close a robot's subtle behavior is to a human.

Third, while *H3* was not fully supported, the results gave an insight into the effect of *proxemics* modality. The condition in which there was least water suggestions taken was in the *minus proxemics* condition. Even though, the state of the art has been focused on using deictic gesture and joint attention gaze for pointing at objects to grab for task-oriented scenarios ([5], [68]), there was no significant difference for these modalities in this study. In addition, the suggestion of object grabbing in this paper was more focused on its social context and implications. It was shown that the participants not only took the object suggested by robot, in this case the water cup, but also drank the water. It may perhaps be linked to the perception of the user to the robot's *situational awareness*, which is the ability to perceive and infer knowledge from the surrounding environment [11]. There is a need for future work to better understand

the potential of proxemics on object manipulation in the shared environment between the user and robot.

Fourth, while *H4a* was validated, *H4b* was not validated and *H4c* was partially supported. The lack of social gestures significantly affected the behavioral alignment for the greeting and closing of the interaction in the *minus gestures* condition. In the greeting part of the interaction, the *minus gestures* condition had significantly less wave gestures performed to greet the robot than in the complete multi-modal and minus social dialogue conditions. For closing waves this difference was even more extreme, with, no wave gestures performed in closing the interaction with the robot in the minus gestures condition. Even further, In Table 5, while the number of participants that did not greet the robot but eventually performed a closing turn at the end of the interaction were highest in the *multi-modal interaction* and *minus gestures* conditions, there was no significant difference. This may also imply that even though there was no behavioral gesture done in the closing of the interaction in the *minus gestures* condition, there was a verbal closing turn.
On the other hand, *H4c* was partially supported. While it was not the *minus gestures* condition that had the least amount of back-channeling alignment performed by users as hypothesized, it was instead the *minus social gaze* condition. This may indicate that gesture mirroring was not the main cause of back-channeling alignment, but rather how naturally the interaction flowed. Conditions with social gaze mechanisms included turn-taking and floor-holding which hold cognitive functions and were accompanied by very short pauses in speech. The users may have performed more back-channeling during these conditions as it was a natural human behavior and as a way to provide the robot feedback that they were in fact still listening to its speech and aligned in the interaction. Thus, the social gaze plays a role in shaping the human-robot interaction seem more instinctive to the human and in forming alignment.

Fifth, a self-reported questionnaire was used to measure liking or 'likability' of the robot and it was hypothesized that the *multi-modal interaction* condition would score higher; however *H5* was not supported. Further behavioral outcomes were annotated and analyzed that might be related to liking of the robot. First, the way the participants addressed the robot was studied. The experiment took place in French with native french speakers and in the french dialect the "you" pronoun is represented by "vous" for formal set-ups and/or with acquaintances and by "tu" for rather informal set-ups and/or with friends. The minus social dialogue condition was significantly different to the minus proxemics condition. The *minus social dialogue* condition was the only one to have users address the robot by its name, e.g., here being "Pepper". In addition, the *minus proxemics* condition had the highest number of participants using informal/friendly pronouns. Further research needs to done to better comprehend what that would signify but it can be concluded at this point that modalities affect the terms participants exercise in the interaction with the robot. Second, during the interaction, the robot is in listening mode at only two phases: at the beginning during the social small talk shown in Table 1 and at the open ended question to measure self-disclosure. At all other times during the interaction, the robot was not in a listening mode and its tablet was required to be used by the participant to answer the questions asked by the robot. However, it was noted that some participants chose to talk while using the tablet and to sometimes justify their choices to the robot and discuss their thought process out loud. While no significant difference was found in the results, the *multi-modal interaction* condition had the highest number of participants who chose to also talk while using the tablet and the *minus gestures* condition had the least. This may give an insight into the effect of *social gestures* on how social participants were with the robot. These findings hint into the type of relationships users formed with the robot based on the multi-modal behaviors they interacted with. There seems to be more to discover and investigate in future works.

## 7 Conclusion

Non-verbal behavior plays a key role in human communication not only by reinforcing and enhancing speech in diverse formats of the interaction, but also by carrying fundamental functions in communication that can stand-alone from speech. However, this non-verbal behavior is not made of only one modality but rather of multi-modalities all composed together to serve their purpose. For this reason, while studying each modality separately may lead to improving human-robot interaction, a deeper understanding of the different modalities when performed together and their combinations as well as their interaction outcomes is imperative for effective use of the multi-modalities of robots in maximizing targeted outcomes. This paper presented work attempting to build such an understanding. The process involved implementing a system of multi-modalities including social gaze mechanisms, different types of gestures, proxemics for navigation in initiating conversations, and social dialogue followed by an evaluation study where participants interacted with the robot in a travel agent scenario. The system and methodology presented in this paper can be as well utilized on other robots. The results showed

various insights into the contributions of modalities in a multi-modal interaction onto several notable behavioral outcomes of the users, including taking physical suggestions, distances maintained during the interaction, wave gestures performed in greeting and closing, back-channeling, how the robot is addressed, and how socially it is treated. It can be concluded that certain modalities in multi-modal behaviors particularly influence the outcomes of the interaction, and at times not in the same way as seen in the state-of-the-art of the modality on its own. Notably, this paper showed how multiple modalities can be combined in an interaction and how subtracting each modality at a time revealed insights about the effect of that modality. For instance, it is now clear how proxemics influence the distances maintained during an interaction and the probability of the user accepting the robot's offer. In addition, social gestures can predict how humans greet the robot and close interactions with it and the utterances the user makes while using other modalities of the robot such as the tablet. Moreover, social gaze shaped how naturally humans back-channel when interacting with the robot but did not have an effect on how much they disclose to the robot. All these findings may lead to further understanding on human-robot interaction and how multi-modal behavior can be used to increase the perceived social intelligence of the robot.

## 8 Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 765955.

The authors would also like to thank the team at INSEAD for assistance with data collection, as well as Hugues Pellerin for advice with the statistical analyses.

## 9 Conflict of interest

The authors declare that they have no conflict of interest.

## References

1. Abele, A.: Functions of gaze in social interaction: Communication and monitoring. Journal of Nonverbal Behavior **10**(2), 83–101 (1986)
2. Admoni, H., Datsikas, C., Scassellati, B.: Speech and gaze conflicts in collaborative human-robot interactions. In: Proceedings of Annual Meeting of the Cognitive Science Society (CogSci '14), vol. 36, pp. 104 – 109 (2014)
3. Admoni, H., Scassellati, B.: Social eye gaze in human-robot interaction: A review. J. Hum.-Robot Interact. **6**(1), 25–63 (2017)
4. Akbıyık, S., Karaduman, A., Goksun, T., Chatterjee, A.: The relationship between co-speech gesture production and macrolinguistic discourse abilities in people with focal brain injury. Neuropsychologia **117** (2018). DOI 10.1016/j.neuropsychologia.2018.06.025
5. Andrist, S., Pejsa, T., Mutlu, B., Gleicher, M.: Designing effective gaze mechanisms for virtual agents. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12, p. 705–714. Association for Computing Machinery, New York, NY, USA (2012). DOI 10.1145/2207676.2207777
6. Andrist, S., Tan, X.Z., Gleicher, M., Mutlu, B.: Conversational gaze aversion for humanlike robots. In: Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction, HRI '14, p. 25–32. Association for Computing Machinery, New York, NY, USA (2014). DOI 10.1145/2559636.2559666
7. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press (1976)
8. Argyle, M., Dean, J.: Eye-contact, distance and affiliation. Sociometry **28**, 289–304 (1965)
9. van Baaren, R.B., Holland, R.W., Kawakami, K., van Knippenberg, A.: Mimicry and prosocial behavior. Psychological Science **15**(1), 71–74 (2004)
10. Birnbaum, G.E., Mizrahi, M., Hoffman, G., Reis, H.T., Finkel, E.J., Sass, O.: Machines as a source of consolation: Robot responsiveness increases human approach behavior and desire for companionship. In: 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 165–172 (2016). DOI 10.1109/HRI.2016.7451748
11. Bolstad, C.A.: Situation awareness: Does it change with age? Proceedings of the Human Factors and Ergonomics Society Annual Meeting **45**(4), 272–276 (2001). DOI 10.1177/154193120104500401
12. Boucher, J.D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., Dominey, P., Ventre-Dominey, J.: I reach faster when i see you look: Gaze effects in human–human and human–robot face-to-face cooperation. Frontiers in Neurorobotics **6**, 3 (2012). DOI 10.3389/fnbot.2012.00003
13. Boucher, J.D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., Dominey, P., Ventre-Dominey, J.: I reach faster when i see you look: Gaze effects in human–human and human–robot face-to-face cooperation. Frontiers in Neurorobotics **6**, 3 (2012). DOI 10.3389/fnbot.2012.00003
14. Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F.: Linguistic alignment between people and computers. Journal of Pragmatics **42**(9), 2355 – 2368 (2010). DOI https://doi.org/10.1016/j.pragma.2009.12.012. How people talk to Robots and Computers
15. Breazeal, C.L., Kidd, C.D., Thomaz, A.L., Hoffman, G., Berlin, M.: Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS pp. 383–388 (2005). DOI 10.1109/IROS.2005.1545011
16. Chartrand, T., Bargh, J.: The chameleon effect: the perception-behavior link and social interaction. Journal of personality and social psychology **76**(6), 893—910 (1999). DOI 10.1037//0022-3514.76.6.893
17. Chiu, C.y., Hong, Y.y., Krauss, R.M.: Gaze direction and fluency in conversational speech. Unpublished manuscript (1995)
18. Clark, H.H., Krych, M.A.: Speaking while monitoring addressees for understanding (2004)
19. Dautenhahn, K., Walters, M., Woods, S., Koay, K., Nehaniv, C., Sisbot, E., Alami, R., Siméon, T.: How may

i serve you? a robot companion approaching a seated person in a helping context. pp. 172–179 (2006). DOI 10.1145/1121241.1121272
20. Delaherche, E., Chetouani, M., Mahdhaoui, A., Saintgeorges, C., Viaux, S., Cohen, D.: Interpersonal synchrony: A survey of evaluation methods across disciplines. IEEE Transactions on Affective Computing **3**, 349–365 (2012). DOI 10.1109/T-AFFC.2012.12
21. Dolinski, D., Nawrat, M., Iza, R.: Dialogue involvement as a social influence technique. Personality and Social Psychology Bulletin **27**, 1395–1406 (2001). DOI 10.1177/01461672012711001
22. Eastwick, P., Gardner, W.: Is it a game? evidence for social influence in the virtual world. Social Influence **4**, 18–32 (2009). DOI 10.1080/15534510802254087
23. Ekman, P., Friesen, W.V.: The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica **1**(1) (1969). DOI 10.1515/semi.1969.1.1.49
24. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Human Factors: The Journal of the Human Factors and Ergonomics Society **37**, 32–64(33) (March 1995). DOI doi:10.1518/001872095779049543
25. Fiore, S.M., Wiltshire, T.J., Lobato, E.J.C., Jentsch, F.G., Huang, W.H., Axelrod, B.: Toward understanding social cues and signals in human – robot interaction : effects of robot gaze and proxemic behavior **4**(November), 1–15 (2013). DOI 10.3389/fpsyg.2013.00859
26. Fong, T., Thorpe, C., Baur, C.: Collaboration, dialogue, human-robot interaction. In: R.A. Jarvis, A. Zelinsky (eds.) Robotics Research, pp. 255–266. Springer Berlin Heidelberg, Berlin, Heidelberg (2003)
27. Grosz, B.J., Sidner, C.L.: Attention, intentions, and the structure of discourse. Computational Linguistics **12**(3), 175–204 (1986)
28. Hall, E., of Congress), C.P.C.L.: The Hidden Dimension. Anchor books. Doubleday (1966)
29. Hall, E., for the Anthropology of Visual Communication, S.: Handbook for proxemic research. Studies in the anthropology of visual communication. Society for the Anthropology of Visual Communication (1974)
30. Hall, E.T.: The silent language / Edward Hall. Doubleday Garden City, N.Y (1959)
31. Hall, E.T.: A system for the notation of proxemic behavior1. American Anthropologist **65**(5), 1003–1026 (1963). DOI 10.1525/aa.1963.65.5.02a00020
32. Ham, J., Bokhorst, R., Cuijpers, R., van der Pol, D., Cabibihan, J.J.: Making Robots Persuasive: The Influence of Combining Persuasive Strategies (Gazing and Gestures) by a Storytelling Robot on Its Persuasive Power. In: Research on Education and Media, vol. 9, pp. 71–83 (2011). DOI 10.1007/978-3-642-25504-5_8
33. Hasson, U., Frith, C.D.: Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. Philosophical Transactions of the Royal Society B: Biological Sciences **371**(1693), 20150366 (2016). DOI 10.1098/rstb.2015.0366
34. Huang, C.M., Mutlu, B.: Modeling and evaluating narrative gestures for humanlike robots (2013). DOI 10.15607/RSS.2013.IX.026
35. Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H., Anzai, Y.: A humanoid robot that pretends to listen to route guidance from a human. Autonomous Robots **22**, 87–100 (2007). DOI 10.1007/s10514-006-9007-6
36. Kendon, A.: Some functions of gaze-direction in social interaction. Acta Psychologica **26**(1), 22–63 (1967)
37. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Studies in Interactional Sociolinguistics. Cambridge University Press (1990)
38. Kennedy, J., Baxter, P., Belpaeme, T.: The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '15, p. 67–74. Association for Computing Machinery, New York, NY, USA (2015). DOI 10.1145/2696454.2696457
39. Kipp, M., Martin, J.C.: Gesture and emotion: Can basic gestural form features discriminate emotions? 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops pp. 1–8 (2009)
40. Kirchner, N., Alempijevic, A., Dissanayake, G.: Nonverbal robot-group interaction using an imitated gaze cue. In: Proceedings of the 6th International Conference on Human-Robot Interaction, HRI '11, p. 497–504. Association for Computing Machinery, New York, NY, USA (2011). DOI 10.1145/1957656.1957824
41. Kong, A.P.H., Law, S.P., Kwan, C., Lai, C., Lam, V.: A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a database of speech and gesture (dosage). Journal of Nonverbal Behavior **39** (2015). DOI 10.1007/s10919-014-0200-6
42. Kruse, T., Kirsch, A., Sisbot, E.A., Alami, R.: Exploiting human cooperation in human-centered robot navigation. In: RO-MAN, pp. 192–197. IEEE (2010)
43. Kucherenko, T.: Data driven non-verbal behavior generation for humanoid robots. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI '18, p. 520–523. Association for Computing Machinery, New York, NY, USA (2018). DOI 10.1145/3242969.3264970
44. Lee, M.K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., Savetsila, S.: Personalization in hri: A longitudinal field experiment. pp. 319–326 (2012). DOI 10.1145/2157689.2157804
45. Leichtmann, B., Nitsch, V.: How much distance do humans keep toward robots? Literature review, meta-analysis, and theoretical considerations on personal space in human-robot interaction. Journal of Environmental Psychology p. 101386 (2020)
46. Leichtmann, B., Nitsch, V.: How much distance do humans keep toward robots? literature review, meta-analysis, and theoretical considerations on personal space in human-robot interaction. Journal of Environmental Psychology **68**, 101386 (2020). DOI https://doi.org/10.1016/j.jenvp.2019.101386
47. Liu, C., Ishi, C.T., Ishiguro, H., Hagita, N.: Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In: Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI '12, p. 285–292. Association for Computing Machinery, New York, NY, USA (2012). DOI 10.1145/2157689.2157797
48. Lucas, G.M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A., Johnson, E., Leuski, A., Nakano, M.: Getting to know each other: The role of social dialogue in recovery from errors in social robots. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI '18, p. 344–351. Association for Computing Machinery, New York, NY, USA (2018). DOI 10.1145/3171221.3171258
49. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago (1992)

50. Mol, L., Krahmer, E., Swerts, M.: Alignment in iconic gestures: Does it make sense? In: B.J. Theobald, R. Harvey (eds.) Proceedings of the eight International Conference on Auditory-Visual Speech Processing (AVSP 2009), pp. 3–8. School of Computing Sciences (2009). Alignment in Iconic Gestures: Does it make sense?
51. Moon, A.J., Troniak, D.M., Gleeson, B., Pan, M.K., Zheng, M., Blumer, B.A., MacLean, K., Crof, E.A.: Meet me where i'm gazing: How shared attention gaze affects human-robot handover timing. In: 2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 334–341 (2014)
52. Mumm, J., Mutlu, B.: Human-robot proxemics: Physical and psychological distancing in human-robot interaction. HRI 2011 - Proceedings of the 6th ACM/IEEE International Conference on Human-Robot Interaction pp. 331–338 (2011). DOI 10.1145/1957656.1957786
53. Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., Ishiguro, H.: Conversational gaze mechanisms for humanlike robots. ACM Transactions on Interactive Intelligent Systems **1**, 12 (2012). DOI 10.1145/2070719.2070725
54. Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., Hagita, N.: Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In: Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction, HRI '09, p. 61–68. Association for Computing Machinery, New York, NY, USA (2009). DOI 10.1145/1514095.1514109
55. Peters, R., Broekens, J., Li, K., Neerincx, M.A.: Robots expressing dominance: Effects of behaviours and modulation. In: 2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 1–7 (2019)
56. Peters, R., Broekens, J., Neerincx, M.A.: Robots educate in style: The effect of context and non-verbal behaviour on children's perceptions of warmth and competence. In: 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), pp. 449–455 (2017). DOI 10.1109/ROMAN.2017.8172341
57. Pickering, M.J., Garrod, S.: Toward a mechanistic psychology of dialogue. Behavioral and Brain Sciences **27**(2), 169–190 (2004). DOI 10.1017/S0140525X04000056
58. Qureshi, A.H., Nakamura, Y., Yoshikawa, Y., Ishiguro, H.: Robot gains social intelligence through multimodal deep reinforcement learning. CoRR **abs/1702.07492** (2017)
59. Saad, E., Neerincx, M., Hindriks, K.: Welcoming robot behaviors for drawing attention. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 368–368. IEEE, United States (2019). DOI 10.1109/HRI.2019.8673325. Video Abstract; 14th Annual ACM/IEEE International Conference on Human-Robot Interaction, HRI 2019 ; Conference date: 11-03-2019 Through 14-03-2019
60. Salem, M., Eyssel, F.A., Rohlfing, K., Kopp, S., Joublin, F.: To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. International Journal of Social Robotics **5**(3), 313–323 (2013). DOI 10.1007/s12369-013-0196-9
61. Sandstrom, G.M., Dunn, E.W.: Social interactions and well-being: The surprising power of weak ties. Personality and Social Psychology Bulletin **40**(7), 910–922 (2014). DOI 10.1177/0146167214529799. PMID: 24769739
62. Schegloff, E.A.: Analyzing single episodes of interaction: an exercise in conversation analysis. Social Psychology Quarterly **50**(2), 101–114 (1987). DOI 10.2307/2786745
63. Schegloff, E.A.: Body torque. Social Research **65**(5), 536–596 (1998)
64. Sciutti, A., Bisio, A., Nori, F., Metta, G., Fadiga, L., Pozzo, T., Sandini, G.: Measuring human-robot interaction through motor resonance. International Journal of Social Robotics (2012). DOI 10.1007/s12369-012-0143-1
65. Shi, C., Shimada, M., Kanda, T., Ishiguro, H., Hagita, N.: Spatial formation model for initiating conversation. In: Spatial Formation Model for Initiating Conversation, Robotics: Science and Systems (2011)
66. Sidner, C.L., Lee, C., Morency, L.P., Forlines, C.: The effect of head-nod recognition in human-robot conversation. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI '06, p. 290–296. Association for Computing Machinery, New York, NY, USA (2006). DOI 10.1145/1121241.1121291
67. Sisbot, E.A., Marin-Urias, L.F., Alami, R., Simeon, T.: A human aware mobile robot motion planner. IEEE Transactions on Robotics **23**(5), 874–883 (2007)
68. Skantze, G., Hjalmarsson, A., Oertel, C.: Turn-taking, feedback and joint attention in situated human-robot interaction. Speech Commun. **65**, 50–66 (2014)
69. Strait, M., Canning, C., Scheutz, M.: Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. In: Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14, pp. 479–486. ACM Press, New York, New York, USA (2014). DOI 10.1145/2559636.2559670
70. Tatarian, K., Chamoux, M., Pandey, A.K., Chetouani, M.: Robot gaze behavior and proxemics to coordinate conversational roles in group interactions. In: 2021 30th IEEE International Conference on Robot Human Interactive Communication (RO-MAN), pp. 1297–1304 (2021). DOI 10.1109/RO-MAN50785.2021.9515550
71. Thorndike, E.L.: Intelligence and its use. Harper's Magazine (140), 227–235 (1920)
72. Torrey, C., Fussell, S.R., Kiesler, S.: How a robot should give advice. In: 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 275–282. IEEE (2013)
73. Travis J. Wiltshire Emilio J. C. Lobato, D.R.G.S.M.F.F.G.J.W.H.H.B.A.: Effects of Robotic Social Cues on Interpersonal Attributions and Assessments of Robot Interaction Behaviors. Proceedings of the Human Factors and Ergonomics Society Annual Meeting **59**(1), 801–805 (2015). DOI 10.1177/1541931215591245
74. Turkstra, L., Ciccia, A., Seaton, C.: Interactive behaviors in adolescent conversation dyads. Language Speech and Hearing Services in Schools **34**, 117–127 (2003). DOI 10.1044/0161-1461(2003/010)
75. Vertegaal, R., Slagter, R., van der Veer, G., Nijholt, A.: Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '01, p. 301–308. Association for Computing Machinery, New York, NY, USA (2001). DOI 10.1145/365024.365119
76. Vinciarelli, A., Pentland, A.S.: New social signals in a new interaction world: the next frontier for social signal processing. IEEE Systems, Man, and Cybernetics Magazine **1**(2), 10–17 (2015). DOI 10.1109/MSMC.2015.2441992
77. Wang, Y., Lucas, G., Khooshabeh, P., De Melo, C., Gratch, J.: Effects of emotional expressions on persuasion. Social Influence **10**(4), 236–249 (2015)
78. Yonezawa, T., Yamazoe, H., Utsumi, A., Abe, S.: Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking.

In: Proceedings of the 9th International Conference on Multimodal Interfaces, ICMI '07, p. 140–145. Association for Computing Machinery, New York, NY, USA (2007). DOI 10.1145/1322192.1322218