

Article

From Social Gaze to Indirect Speech Constructions: How to Induce the Impression That Your Companion Robot Is a Conscious Creature

Boris M. Velichkovsky ^{1,2,3,*}, Artemiy Kotov ^{1,3,4} , Nikita Arinkin ^{1,3}, Liudmila Zaidelman ^{1,3}, Anna Zinina ^{1,3,4,*} and Kirill Kivva ^{3,5}

- ¹ National Research Center “Kurchatov Institute”, 123182 Moscow, Russia; kotov_aa@nrcki.ru (A.K.); arinkin_na@nrcki.ru (N.A.); laburnum-yz@yandex.ru (L.Z.)
 - ² Faculty of Psychology, Technische Universitaet Dresden, 01062 Dresden, Germany
 - ³ Scientific Department, Scientific Center of Cognitive Programs and Technologies, Russian State University for the Humanities, 125047 Moscow, Russia
 - ⁴ Laboratory of Cognitive Research of the Foundations of Communication, Moscow State Linguistic University, 1190034 Moscow, Russia
 - ⁵ Computer Science & Control Systems Faculty, Bauman Moscow State Technical University, 105005 Moscow, Russia; k.kivva@gmail.com
- * Correspondence: velch@applied-cognition.org (B.M.V.); anna.zinina.22@gmail.com (A.Z.)

Featured Application: To ensure our fluent and partner interaction with the new generation of interactive artificial agents, these agents must use intermodal communication channels, including eye movements and diverse behavioral manifestations of intellectual and emotional attitudes towards the referents, denoted in communication. We intended to implement an imitation of the so-called social gaze, as it is of the primary significance for human–robot interactions. The suggested model can be used in the next generation of companion robots to maintain a feeling of eye contact and natural personal interaction. Such robots may include personal companions, smart home interfaces, information assistants in public transport, offices, or shops, toys, and educational robots.

Abstract: We implemented different modes of social gaze behavior in our companion robot, F-2, to evaluate the impression of the gaze behaviors on humans in three symmetric communicative situations: (a) the robot telling a story, (b) the person telling a story to the robot, and (c) both parties communicating about objects in the real world while solving a Tangram puzzle. In all the situations the robot localized the human’s eyes and directed its gaze between the human, the environment, and the object of interest in the problem space (if it existed). We examined the balance between different gaze directions as the novel key element to maintaining a feeling of social connection with the robot in humans. We extended the computer model of the robot in order to simulate realistic gaze behavior in the robot and create the impression of the robot changing its internal cognitive states. Other novel results include the implicit, rather than explicit, character of the robot gaze perception for many of our subjects and the role of individual differences, especially the level of emotional intelligence, in terms of human sensitivity to the robotic gaze. Therefore, in this study, we used an iterative approach, extending the applied cognitive architecture in order to simulate the balance between different behavioral reactions and to test it in the experiments. In such a way, we came to a description of the key behavioral cues that suggest to a person that the particular robot can be perceived as an emotional and even conscious creature.

Keywords: human–robot interaction; social gaze; eye-to-eye contact; emotional interfaces; eye–brain–computer interfaces; attention; reflection; usability; brain hemispheric lateralization



Citation: Velichkovsky, B.M.; Kotov, A.; Arinkin, N.; Zaidelman, L.; Zinina, A.; Kivva, K. From Social Gaze to Indirect Speech Constructions: How to Induce the Impression That Your Companion Robot Is a Conscious Creature. *Appl. Sci.* **2021**, *11*, 10255. <https://doi.org/10.3390/app112110255>

Academic Editor: Attila Kovari

Received: 20 August 2021

Accepted: 22 October 2021

Published: 1 November 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Gaze and speech are the primary cues of attention and intellect of another person. Within personal communication, gaze is the first communicative sign that we encounter, and which conveys information about the mental state, attention, and attitude of our counterpart. The direction of the opponent's gaze is easily detected by another person (and measured by conventional eye trackers), so gaze not only has a perceptive function but a significant communicative effect as well, which is extensively described in the literature on human–human [1,2] and human–robot [3,4] interactions. Overall, behavioral patterns of the eyes, including the eyelids and brows, can be studied and modeled as a symptom of the internal cognitive operations of the subject, or as a communicative cue, exposing to the addressee the internal state of the subject, or intentionally conveying meanings in communication [5]. We examined the perception of gaze behavior for short periods of direct eye contact, as well as for longer durations of human–robot interactions, in which the robot had to change its gaze direction between several objects of interest in the environment. In this work, we used an iterative approach, trying (a) to extend the robot control architecture in order to simulate gaze dynamics, (b) to test the architecture in experiments within typical communicative situations, and (c) to observe the key features—communicative cues or strategies—that have to be included in applied and theoretical architecture in order to make robots' behavior more reflective from the point of view of a human. Additionally, we considered the interaction of eye movements with other activities that evolve within the framework of continuous meaningful interactions. Not immediately obvious responses, such as indirect speech acts, may be of particular importance for inducing the feeling that your partner is not an automat.

As studied from the point of view of a human subject, vision and, therefore, gaze are the most important channels of perception. Gaze is directed by the attention system and can be used in humans and in primates in the ambient mode (for orientation in the space) or in the focal mode (to examine a particular object of interest) [6,7]. During a simulation, it is important that ambient and focal systems compete to gain control over the direction of the eyes, as attention to different objects of interest may raise the internal competition to shift the gaze to each of the objects of attention. This creates a bottleneck in a situation of time pressure, and the gaze system becomes a limited resource that must meet the numerous requirements of the attention and communication systems—to direct the attention to each interesting object, to simulate communicative behavioral patterns, and to blink [1,3]. The units to control the gaze can experience not only positive activation—arousal—but also negative influence—suppression. Although the addressee is a natural object of interest and, thus, must attract the attention, a long and direct gaze constitutes a face threatening act [8] in many human cultures and should be limited in terms of the communicative theory of politeness. In this way, a separate unit should withdraw the gaze from the addressee after some critical time periods.

The eyelids and brows system, controlled by musculi or action units, according to the Facial Action Coding System [9], may also serve the attention system by squinting or widely opening the eyes, as well as expressing numerous cognitive and emotional states. On one hand, this behavior can be out of the subject's voluntary control, but the corresponding superficial cues can still convey information to the addressee, thus serving as normal signs in communication. On the other hand, the capacity of voluntary control over the system of the gaze/lids/brows allows the subject to express intended meanings; one may look at an object to designate it, or intentionally frown to express the concern. The variable degree of intentionality within the control of communicative cues is described as Kendon's continuum [10,11], where the unintentional and uncontrolled behavioral patterns stay at one end, while the controlled nonverbal signs, quite like the signs of natural language, are located at the other end. Departing from this point, the expressive possibilities of the gaze system are extensively studied within the theory of communication; a comprehensive review can be found in [12]. Gaze is also an important feature in the natural interfaces of

companion robots, and thus, it is widely implemented and evaluated within applied robotic systems, for example, in the robots Cog [13], Kismet [14], and Infanoid [15], and others.

The attention of researchers and developers to artificial gaze systems is attracted by various features of the natural gaze.

- (a) **Gaze direction** is the primary object of study, and in applied systems, the direction of gaze is critical for effective interaction. At the same time, the perception of gaze is quite complex and may include the evaluation of the position of the eyes relative to the rotation of the head. For example, the gaze direction of a back-projected robot is perceived more naturally and precisely if its eyes are moving relative to the head and not fixed in the central position [16].
- (b) Saccades, pupil dilatation, and ocular torsion constitute the micromovements within the gaze system. They are the main objects of studies regarding the psychology of attention but are not widely implemented in artificial interfaces. An extensive study on gaze micro-features, including eye saccades, eyelid micro-movements, and blinking, and their relation to the level of trust that they arouse in users, is represented in [17].
- (c) Movements of eyelids and brows and blinking. Most virtual agents and even some physical robots are designed to blink and move their eyelids and eyebrows. In [18], the behavioral modes of a physical robot were compared, including non-blinking and blinking according to statistical models, and blinking according to physiological data. According to this study, a robot with a physiological blink seemed more intelligent to the respondents. In [19], the researchers varied the number of blinks performed by the agent. It was determined that respondents have a feeling that the agent is looking at them when the number of blinks of the agent exceeds the number of blinks of the respondent. Thus, eyebrows and eyelids combined can express numerous expressive and emotional patterns, which are widely studied in the field of the psychology of emotion and are also used within the existing interfaces of Kismet, Mertz, iCat, and others.
- (d) Responsive gaze is studied as the ability to reply to another's gaze by looking at the opponent and maintaining the balance of gazes during communication. This topic is extensively studied by Yoshikawa and his colleagues with regard to human–robot interactions. In experimental studies, the researchers showed that the robot that responded with its gaze to the gaze of the interlocutor provided the subjects with the feeling of gaze contact and was evaluated as a more positive counterpart [19,20].
- (e) Joint attention is studied as the ability to concentrate on the object of interest of another person. A robot's ability to support joint attention with humans is considered an important feature for companion robots, as it makes them look more competent and socially interactive while solving spatial tasks with humans [21]. In a study from our lab, subjects learned to control a robotic device using the joint attention gaze patterns. The robot was prompted by a responsive gaze of a human and then followed the direction of the human gaze, simulating joint attention to select the location for the required action of the robot [22].

These studies on the superficial expressive patterns of gaze are combined with a major area, in which human gaze is used as a controller in interfaces to position a cursor, select objects, or navigate in the environment. Although one can hardly underestimate the importance of these studies, in this article, we want to address the communicative aspect of the gaze—the ability of the eyes to express the internal cognitive processes of an agent. This feature, if simulated by artificial architectures, can give us a better understanding of the dynamics of a human gaze in its connection to internal cognitive states, and when applied, can make companion robots look more intelligent and attractive. Such an interface can act not as an extension of a human body, but as a companion, utilizing its gaze to report to a human its internal or communicative states. In this respect, we extended an applied cognitive architecture to model gaze behavior for an experimental companion robot, F-2, and tested the model in two situations of storytelling as well as a collaborative game situation. We used gaze direction to balance looks at the addressee, the object of

interest (game pieces), or side gazes. For all the experiments, we expected that the robot, controlled by the balance model, would look more attractive and/or intelligent than the robot controlled by a simpler direct gaze management system. Within the experiments, we also applied some movements of the eyelids and eyebrows to make the gazes more vivid. We tested the models of joint attention and responsive gazes in the collaborative game and storytelling. At the same time, we did not apply micromovements such as microsaccades, drifts, pupil dilatation, and ocular torsion because their influence on the impression of the robot was discovered to be insignificant during the preliminary tests.

2. Methods and Architecture of the Model

In our study, we relied on an iterative approach, including the basis of a multimodal corpus and previous experiments from which we selected communicative functions that might distinguish conscious behavior if simulated by a companion robot. We tried to extend the previously developed architecture of a companion robot to simulate the observed functions and attempted to combine them with other, previously developed behavioral cues and strategies. Furthermore, we tested the impact of the modeled cues within the experiments to evaluate their contribution to the perception of the robot as a conscious creature. In this series of experiments, we extended the model to simulate the gaze management patterns in three symmetrical communicative situations.

In previous experiments, we have noticed that people may follow some nonverbal cues of the robot and attribute internal states to some specific movements and to the changes of the robot's gaze direction. In the developed architecture, nonverbal cues are combined with verbal responses; a robot can perform gestures and/or utterances in response to different incoming stimuli, including users' utterances, tactile events, and users' movements. In order for the robot to react to the most essential stimuli, or to express the most essential internal states, we implemented a compound architecture in which the internal units compete and concurrently gain control over all or some of the robot's effectors to perform movements or phrases. This architecture is also applicable to the control of gaze direction; an applied gaze management system should be organized to hold and constantly solve the conflicts between numerous units of attention and expression, competing to gain control over the direction of eyes. M. Minsky [23] suggested the classic architecture of proto-specialists to handle such conflicts in a system, by which numerous cognitive or emotional units try to control the body of an artificial agent or robot. Following Minsky, each proto-specialist constitutes a simple cognitive unit, responsible for handling some simple stimulus (e.g., a threat) or goal (e.g., hunger). Proto-specialists change their activation over time; the leading unit gains control over the required effectors, and then discharge, being satisfied, while losing the initiative in favor of other proto-specialists. This architecture has been extended by A. Sloman within the Cognition and Affects Project (CogAff). He has distinguished reactive units into different cognitive levels; while emotions and drives stay at the primary level, deliberative mechanisms (second level) or reflective reasoning (third level) can also gain control over the body of the agent to suggest the execution of longer and more sophisticated behavioral programs. In this architecture, deliberative reasoning may suppress emotions, conciliating the agent via rational reasoning. On the other hand, emotions can return the reasoning process of the agent to the object of desire or anxiety, thus reactivating themselves via the mechanism of positive feedback [24].

Our companion robot, F-2, has quite simple hardware and is controlled by an extended software architecture, which processes natural language texts (oral speech and written sources), as well as the events from a computer vision system to provide the robot with reasonable reactions (see Figure 1). Its central component conceptually corresponds to the CogAff architecture and operates with a set of scripts that are activated by incoming events and sending their behavioral packages to be executed by the robot.

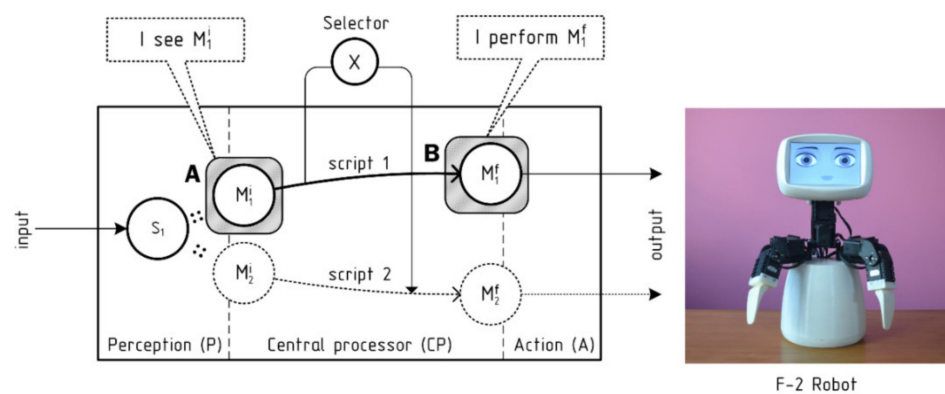


Figure 1. The general processing architecture of the F-2 robot.

Different incoming stimuli are processed within several components. Speech and text are processed by a semantic parser [25], which receives written text at its input, passes the stages of morphological and syntactic analysis, and constructs a semantic predication—a frame—for each sentence. For oral messages, an external speech-to-text service can be used to convert the signal to the written form (in our case, *Yandex speech API*). Several recognition variants can be processed in parallel in case they are generated by the recognition service. The ability to operate with semantic predications (frames) is the key feature of the suggested model and its main difference from the CogAff architecture. We used semantic predications in a form, which is usual for the representation of a sentence meaning in linguistics [26,27]. Each frame is divided by valencies, including the predicate, agent, patient, instrument, time, location, and so on. We used a list of 22 valencies, following the method by Fillmore [28]. Each valency within a frame is represented by a set of semantic markers—nuclear semantic units extracted from the meanings of words within this valency or assigned by the corresponding visual processing component. We used a list of 4835 semantic markers designed on the basis of (a) the list of semantic primes by A. Wierzbicka [29], (b) categories of the semantic dictionary [30], and (c) two-level clustering of word2vec semantic representations. Within the annotation (c), one marker is assigned to each word within a cluster, so two markers are assigned to each word in a two-level clustering tree [31]. The method was designed to handle automatic clustering with a tree of arbitrary depth. A two-level tree is presently used for the approbation of the approach.

Visual stimuli are processed by dedicated software components, and each of these also produces semantic predications to be processed by the central component of the robot—scripts. The robot keeps and updates the coordinates of recognized objects and events in 3D space, and generates semantic predications for the script component, such as “someone is looking at me”. Face detection and orientation are processed with the help of the OpenCV library. The recognition of tangram puzzle elements is carried out with a specially developed tool using the *IGD* marker system. Each movement of a game piece is also converted to a semantic predication (such as “the user correctly moves game piece No 7”), and accounted for in the module, which monitors the progress in the solution of the puzzle. The extended processing architecture, with examples of semantic predications for different stimuli, is presented in Figure 2.

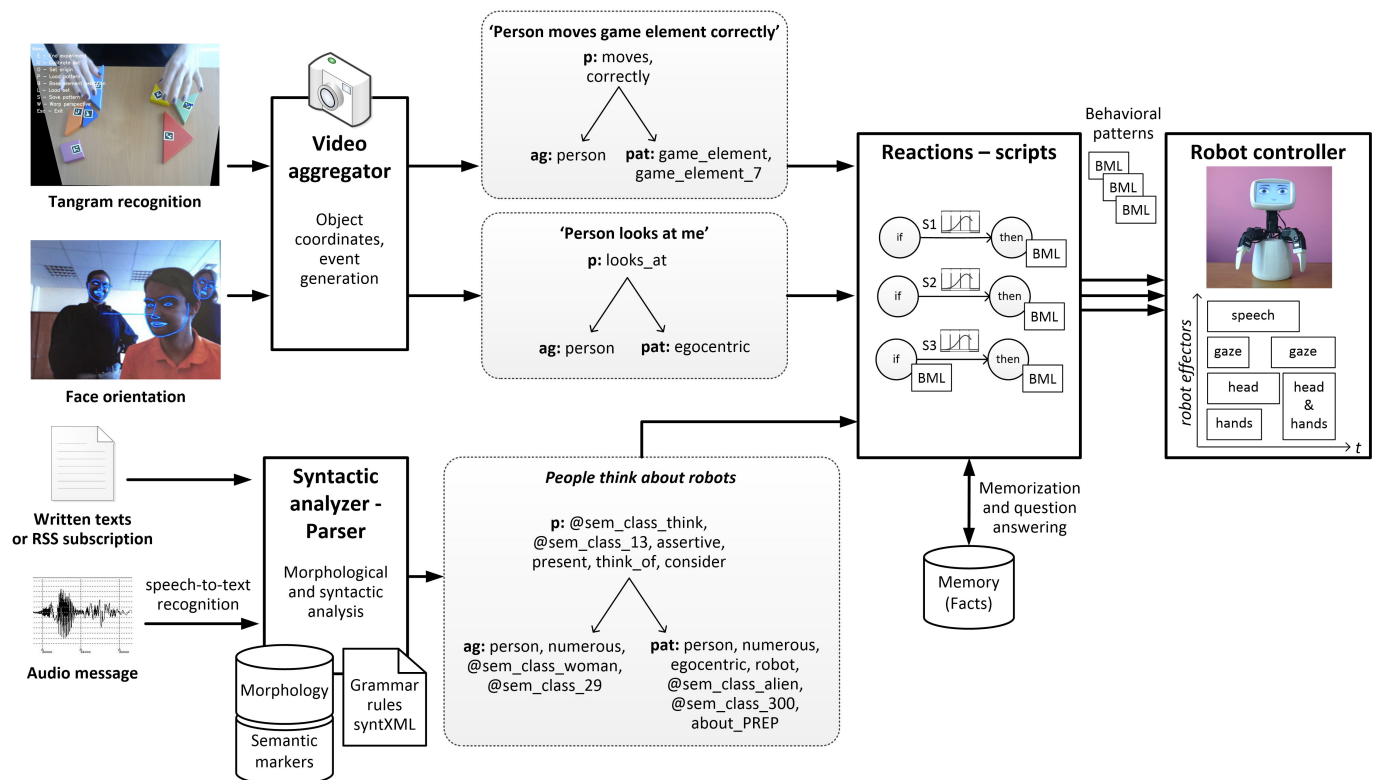


Figure 2. The extended processing architecture of the F-2 robot.

Within the suggested architecture, we used a list of scripts to process the stimuli and execute responsive reactions, as well as to support the competition between scripts in order to establish compound behavior (see further [25]). Each script is a type of production—an if-then operator. Its premise (if condition) and corollary (then condition) are also defined as semantic predications, so an input stimulus may invoke a script, which may, in turn, create a new semantic predication to invoke further scripts. A script may contain behavioral patterns to be executed once the script is activated so it can control the effectors of the robot, including the gaze direction. As the semantics of input sentences and real-world events are represented in a unified way, the system may (a) receive new knowledge from a visual observation or a text description, and (b) react to each incoming representation—such as an incoming user’s gaze, a user’s move within a game, or a user’s utterance—and balance the reactions to each stimulus, regardless of its modality. It can also handle the competition between numerous scripts of the same type, for instance, when a user makes many moves within the game or when many people look at the robot simultaneously. For each incoming stimulus, we calculated its similarity with all the available premises of scripts via a modified Jaccard similarity coefficient to evaluate the number of semantic markers expected by a script premise and present in the stimulus. The best script is selected and activated proportionally to its similarity with the stimulus and its sensitivity, for example, its prior activation by preceding stimuli. Each activated script sends the desired behavioral pattern to the robot controller. The controller monitors the available effectors of the robot and the list of behavioral patterns suggested for executions by all the activated scripts; further, it executes a pattern from the most activated script as soon as the required effectors are available. A script loses its activation either immediately, when its behavioral patterns are executed on the robot, or gradually in time, while it waits for the expression and becomes irrelevant.

In this architecture, visual stimuli, as well as stimuli of other modalities, are naturally filtered by the agent depending on their subjective relevance. A relevant stimulus better matches scenarios and forces their activation, thus ensuring that the robot will respond to this stimulus with its behavior. Less relevant stimuli cause the moderate activation

of scripts that can be expressed in the absence of a more relevant stimulus, or otherwise decelerate in time. Irrelevant stimuli can match no script and not get processed at all. This architecture allows the agent to respond to the stimuli, following their subjective relevance to the agent and depending on the available resources—most notably, time.

In our script component, we used 82 scripts for emotional processing and 3500 scripts for the rational processing of events. Rational scripts allow for the resolution of speech ambiguity and provide the foundation for the planned mechanism of rational inference. The most relevant script for an input event is selected based on the number of semantic markers of the script premise present within the input event, and it is calculated via the modified Jaccard similarity coefficient. For each stimulus, a separate instance of the script is created with the calculated activation. For example, if several people look at the robot, then a separate script instance with a responsive gaze behavioral pattern is created for each person. Each script is linked to a behavioral pattern, defined with the Behavior Markup Language (BML) [32,33]. Specific movements, such as gestures and head and eye movements, that are not attached to the coordinates of the surrounding objects, are selected from the REC emotional corpus [34], which was designed in Blender 3D rendering software and stored in the LiteDB database as the arrays of coordinates for each of the robot's effectors in time.

The materials, methods, human participants, and the results of three actual experiments with robots, which were developed on the basis of the architecture, are described in detail in Section 3.

3. Experimental Studies and Results

We consecutively applied the model to three major communicative situations, in which (a) the robot tells a human a story, so the denotatum is represented by the robot; (b) the human has to solve a special puzzle and the robot follows the solution and gives advice—here the denotatum is explicitly represented to both the human and the robot; (c) the robot listens to a story that is being told by the human, and therefore, the denotatum here is represented by the human. In all these conditions, the robot has to apply different gaze control modes to balance between the social (responsive) gaze, the side gaze, and the gaze to the physical object in problem space, if it exists. Although natural gaze behavior is rather compound and may differ in varying communicative situations, we iteratively extended the suggested model in order to cover these conceptually symmetric types of communicative situations.

In general, eye movements tend to have at least an implicit influence on the robot's attractiveness to a user. In our recent study [35], we investigated the contribution of the robot's effectors' movements—such as those of the hands, head, eyes, and mouth—to their attractiveness to the user. The subjects were students, teachers, and counselors of an educational camp ($n = 29$, 17 females, mean age of 19). During the experiment, the subjects were asked to listen to five short stories narrated by the robot. The subjects had to listen to each story twice in random order—with the robot using all active organs—*full-motion mode*, and without movements of a specific active organ (no movement of the eyes, head, hands, and no mouth animation)—*deficit mode*. After presenting two experimental conditions for each story, the subject chose their most preferred type of the robot's behavior. The subjects were also asked to describe the difference between the two modes of storytelling by the robot. As expected, the subjects in all cases chose the robot that used full-motion behavior. The experimental study revealed that users are more likely ($p < 0.01$, Mann–Whitney U-test) to prefer a robot that uses gestures, head movements, eyes, and mouth animation in its behavior compared to a robot for which some part of its body is stationary. The impact of the robot's eye movements on the user was quite implicit; the subjects significantly more often ($p < 0.05$, Mann–Whitney U-test) preferred the robot with eye movements to the robot with motionless eyes, but they seldom ($p < 0.05$, Mann–Whitney U-test) explicitly noticed the difference between these two modes. The implicit nature of the preference was supported by the verbal responses to a subsequent questionnaire; the subjects rather

indicated the irrelevant differences between the robots, for example, they assumed that one of the robots spoke faster, said something wrong, or was in some way kinder and/or more interested.

3.1. Experiment 1: The Robot Tells a Story

To reproduce the pattern of complex gaze behavior on the robot, we developed a model in Python in which different scripts change their activation in time, and the script with the highest activation takes control of the robot's eyes. Figure 3 shows the functioning of the model with three scripts: (a) to think—this state changes sinusoidally over time; (b) to speak—the model received data about the starting time of the utterance and the phrase accent; (c) to pay attention to the person—this state increases linearly if the robot was not looking at the person and resets its activation as soon as the robot looks at the person. Every 40 milliseconds, the system calculates the leading script and allows it to set the direction of gaze. This model was used in one of the experimental conditions and was compared to some simple models of gaze management, which are described later. The eyelids and mouth did not move in this experiment. The robot detected the position of the human face within the environment and directed its gaze at the location of the human's eyes so that the subjects could freely move in the space in front of the robot.

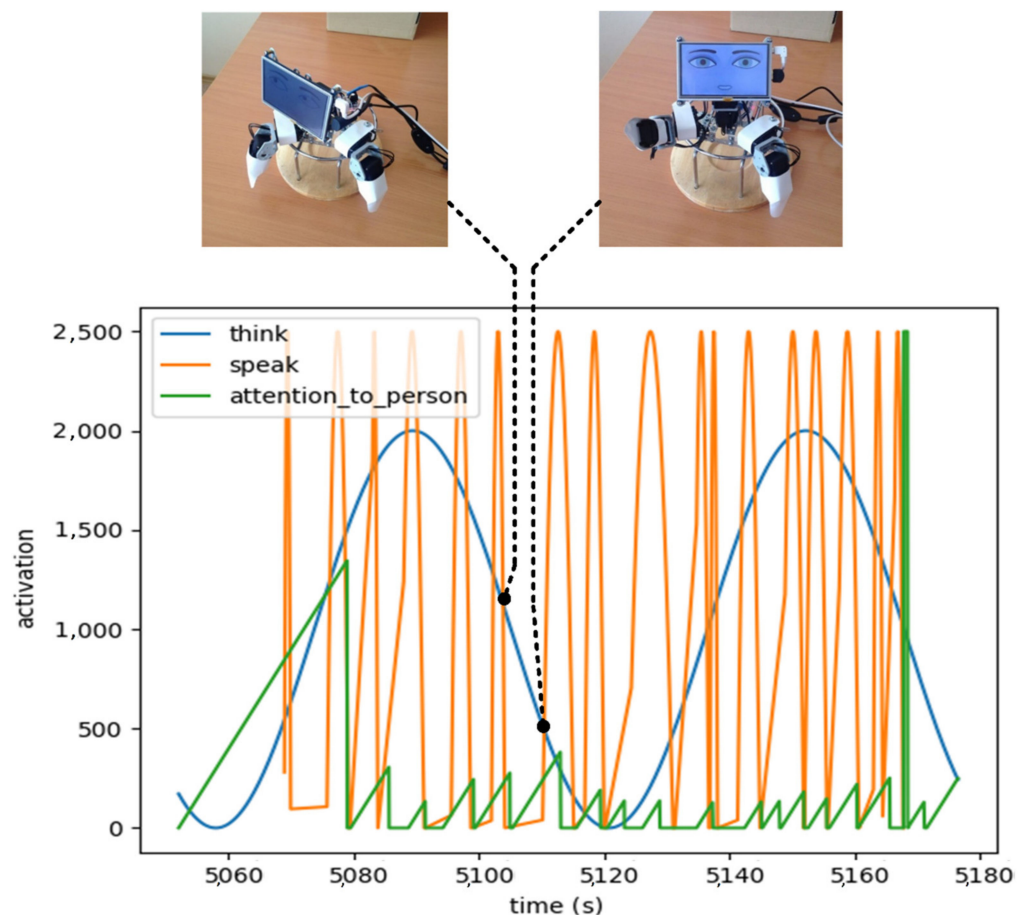


Figure 3. Gaze control model via the competition of control states; gaze direction is controlled by the leading state.

In the experiment, the participants ($n = 14$, six females, mean age of 27.2 years) evaluated the contribution of the gaze model to the robot's attractiveness (a preliminary report was given in a recent paper [36]). During the experiment, the robot narrated eight short stories. Each participant had to evaluate the robot on two modes of behavior (or conditions) according to Osgood's semantic differential scales. In Condition 1, the robot

switched its gaze direction every 6 s from left to right, in turn, not looking at the user. In Condition 2, the gaze direction was controlled by the leading script, as mentioned above, and the robot could direct its gaze to the user. The activation of each of these scripts varied in time. When the leading state was changed, the new leader changed the gaze direction of the robot, as shown in Figure 3. The robot's utterances were accompanied by iconic gestures, which were similar in the two experimental conditions.

Results and a Preliminary Discussion

In the experimental setting, the subjects significantly more often ($p < 0.01$, Chi-Square) preferred the robot controlled by the balancing gaze model (Condition 2). This robot was significantly more often described as friendly, attractive, calm, emotional, and attentive ($p < 0.01$, Chi-Square). The results allow us to evaluate the contribution of the gaze direction to the positive perception of the robot and prove the effectiveness of the developed model of gaze behavior in the situation of human–machine interactions.

3.2. Experiment 2: The Robot Helps Humans in a Game

To further evaluate the influence of social gaze, we tested the model in a situation where the robot was helping a human solve a tangram puzzle. The goal of the experiment was to evaluate the influence of oriented gestures with directed gaze on a human ($n = 31$, 12 females, mean age of 27.4). The task of the participants was to arrange the elements within a given contour on a white sheet. During the experiment, the participant was to complete several figures, including a parallelogram, a fish, a triangle, and a ship. For each task, the game elements were placed in front of the participant on the left and right sides of the playing field. Two paired elements (two large triangles and two small triangles) were always placed on different sides of the playing field. The robot helped the human, indicating in speech which game element to take and where to place it (Figure 4). In half of the tasks, the robot used oriented communicative actions (pointing hand gestures, head movements, and gaze) to indicate the required game element, and then the correct position to place it (Condition 1). For the paired elements, the left or right element was chosen randomly. In the other half of the tasks, the robot used non-oriented, symmetric gestures while providing the same speech instruction (Condition 2). In the first condition, the direction of the robot's gaze was considered an indication. In its speech instructions, the robot always referred to an element by its shape and size, not by color. So, when the robot named paired elements, the reference was ambiguous and could indicate the element on either the right or the left side of the playing field.

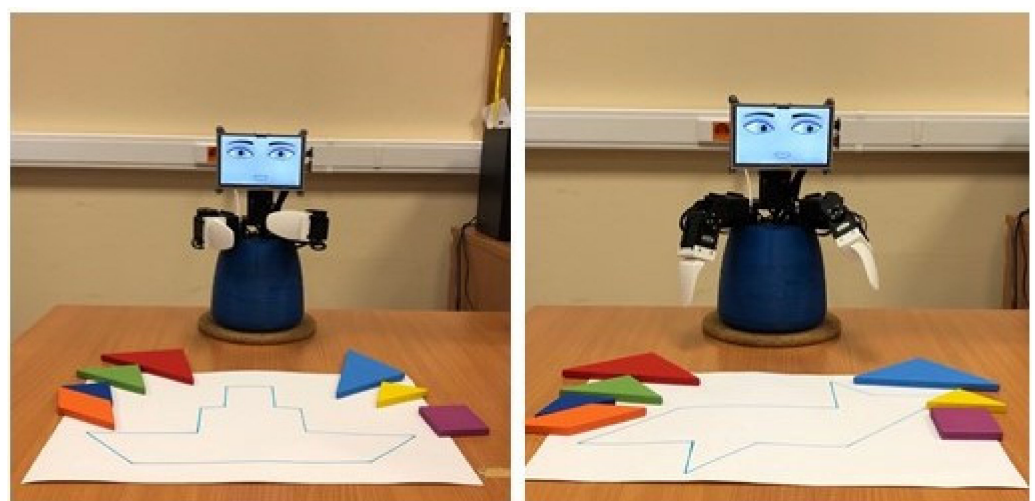


Figure 4. Robot refers to the game element with speech (with ambiguous reference) or with speech and gesture.

Results and a Preliminary Discussion

According to the results of the experiment, most of the participants (64.5%) significantly ($p < 0.01$, Mann–Whitney U-test) preferred the robot pointing to the tangram elements and their locations with eyes and hands. At the same time, the difference between both experimental conditions was not obvious to the players; only half of them (15 people; 48.4% of the total group) noticed the difference between the robot's performance with and without the pointing gestures and gaze. Therefore, the evaluation of the robot can again be implicit because the participants did not distinguish between the two experimental conditions verbally, but simultaneously preferred the robot that indicated the necessary element and its location in the contour with the help of its head, eyes, and hand movements. This experiment is described in more detail elsewhere [37].

The results clearly show the importance of directional gaze for communication between a robot and a user. Directional gaze is perceived as a reflection of the robot's internal state. In the speech instructions, the robot prompts the user, and gaze is used to provide the reference for this prompt. It is this correspondence of mimicked expression to the robot's internal intention that is positively perceived by the subjects; however, it occurs at the implicit level.

3.3. Experiments 3: The Robot Listens to a Story

One of the key characteristics of social gaze is the ability to change the gaze direction following the addressee. To simulate the responsive gaze of the robot, we decided to examine a situation in which a person tells the robot a story and the robot acts as an active listener by demonstrating different modes of responsive gaze. The purpose of the new experiment was to study the effect of a robot's responsive gaze on its attractiveness to the user. In this experiment, two robots reacted to the communicative actions of a human in two different modes. In one case, the robot directed its gaze at the user; in the second case, it demonstrated side gaze, which is typical for the condition of thoughtfulness or when regulated by a politeness strategy to avoid face threatening acts aimed at the addressee [8,38]—video is provided within the Supplementary Materials. Within the experiment, subjects had to tell two robots stories following a list of pictures by Herluf Bidstrup (Figure 5; see more about his artwork at https://en.wikipedia.org/wiki/Herluf_Bidstrup, accessed on 1 October 2021). Each picture was represented as a stack of cards in random order. A total set of six stories was used; each subject had to tell three stories to each robot, switching to the other robot after each story.

A total of 46 subjects participated in the experiment (mean age of 27 years, 33 females). Before the experiment, the F-2 robots were introduced to the subjects. The respondents were told that the developers were currently trying to train the robot to follow the story narrated by a human. The main purpose of the study, which was to investigate the effect of the robot's responsive gaze, was not included in the introduction. We examined the following hypotheses: (a) the robot is perceived as more attractive if it establishes gaze contact with the user; (b) respondents with a high level of emotional intelligence better distinguish the behavioral patterns of the robots.

The orientation of the user's attention was roughly identified automatically by the orientation of the user's face and was recognized by a computer vision component based on OpenCV. This system was chosen as a possible "built-in" solution for emotional companion robots, allowing us to avoid the calibration typical for eye-tracker experiments, and thus, maintain more natural communication with the robots. To find a vector of the human face orientation that we interpret as a gaze vector, we built an approximate 3D model of human facial landmarks and developed the following steps. Human faces are detected within the camera video frames using a linear SVM classifier based on an advanced version of HoG features [39] implemented in the Dlib library. The image intensity histogram is normalized inside a box bounding the face. Then, landmarks of the found face are detected using an Ensemble of Regression Trees approach described in [40] and implemented in the Dlib library. These 2D landmarks' positions and 3D coordinates of the corresponding model

points are used to solve a perspective-n-point (PnP) problem using the method suggested in [41] an implemented in the OpenCV library. Solving the PnP problem provides the orientation of the human face in 3D with respect to the camera. Next, a 3D gaze vector is built starting between the human's eyes and using the calculated face orientation, and the vector's coordinates are stabilized using a simple Kalman filter, assuming these coordinates to be independent. Then, a constant offset is added to the gaze vector's endpoint to make the vector point at the camera when the human is looking directly at the robot. The angle between the gaze vector and the direction from the camera center to the human's nose bridge provides information on if the human is looking at the robot.






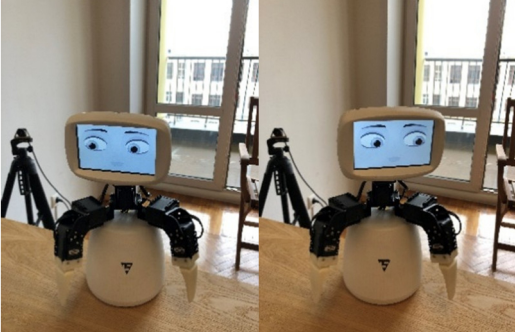
Figure 5. Example of stimulus material—a graphic story by Herluf Bidstrup. Each story was represented as a stack of separate cards in random order.

The procedure of the experiment was as follows. A participant sat in a room in front of the two robots, identified by square and triangle marks on their bodies. The first robot said that it is ready to listen to a story. The subject took the cards from the stack and had to organize them into a story. While narrating, the subject was asked to show the robot the card corresponding to the current part of the story. At the end of the story, the robot expressed its gratitude and asked to tell the next story to the other robot. Within the setup of the experiment, three zones of attention were distinguished—a participant can organize the cards on the table and can show the cards to the left or the right robot. This setup ensured higher precision of the user recognition system, which was detecting the face orientation rather than the gaze itself. The subject could communicate with the left and right robots from the same position at the table, so the robots constantly maintained the corresponding gaze behavior. For example, each robot could respond to the user's gaze even if the user told a story to the other robot. Following the experiment, the subject had to express their preference for the triangle robot, the square robot, or both robots equally, as well as evaluate each robot on a five-point scale.

While “listening”, the hands of the robots were controlled by the inactive component, which initiates permanent minor movements. Additionally, during the speech replies, the hands were controlled by more significant gestures that were coordinated to the utterances. At the same time, the head and eyes were constantly controlled by the two scripts responsible for the social gaze. When a person looked away from the robot (e.g.,

looked aside or looked at the table to follow the cards), each of the robots looked down—at the table with the cards. When a person looked at the robot, the first robot (marked with a square) looked back, raised its head, opened its eyelids, and raised its eyebrows (Table 1), while the second robot (marked with a triangle) demonstrated the side gaze, looking left or right and randomly changing the direction.

Table 1. Robots' reactions to human gaze: gaze responsive and gaze aversive behavior.

Title 1	Before or After the User's Gaze	During the User's Gaze
Gaze-responsive robot (marked with a square)		
Gaze-averse robot (marked with a triangle)		

Left- or right-side gazes are selected randomly

We used the Emotional Intelligence Test (EmIn) [42,43] to evaluate the level of emotional intelligence of the participants. The questionnaire is based on the interpretation of emotional intelligence as the ability to understand one's own and others' emotions, as well as to control one's own emotions. This understanding implies that a person (a) recognizes the presence of an emotional experience in oneself or another person, (b) identifies the expression and can find a verbal designation for one's own or another's emotion, and (c) understands the causes and effects of that emotional experience. Controlling emotions means that the person controls the intensity and external expression and can, if necessary, arbitrarily evoke a particular emotion. The test suggests an aggregated score of emotional intelligence, but we expected that the recognition of others' emotions would play a major role in the perception of the robots.

We also evaluated the accuracy of the recognition system. The automatically obtained data was compared with the observed gaze of the subjects during the experiment. The subject's gaze on the robot was defined as a movement of the gaze and head in the direction of the robot starting from 1 s. The number of reactions roughly corresponded to the number of gazes. The system was performing with a redundancy of 133% (for 1988 gazes, it demonstrated about 2650 responsive actions), the majority of redundant reactions were demonstrated during long human gazes (more than 30 s), when the system could execute several gaze responses. The statistics of real gazes were obtained by qualitative analysis of video recordings of the experiment. Taking into account the slight redundancy in performance for long gazes, the system was considered appropriate to execute gaze control in the actual communication.

Results and a Preliminary Discussion

According to the results of this experiment, if a person explicitly preferred a specific robot or evaluated it higher, they were assigned to the corresponding preference group. Within these groups, 26% of people ($n = 12$) preferred the gaze-avoiding robot (triangle), 28% of people ($n = 13$) preferred the gaze-responsive robot (square), and 46% ($n = 21$) evaluated the robots as equal. The subjects who preferred a specific robot demonstrated significantly different results on the scale of recognition of the emotions of others within the EmIn test. People who preferred the gaze-responsive robot had a level of emotional intelligence significantly higher than the group who preferred the gaze-avoiding robot (Figure 6). The subjects did not demonstrate any significant difference in any other scale of the emotional intelligence test.

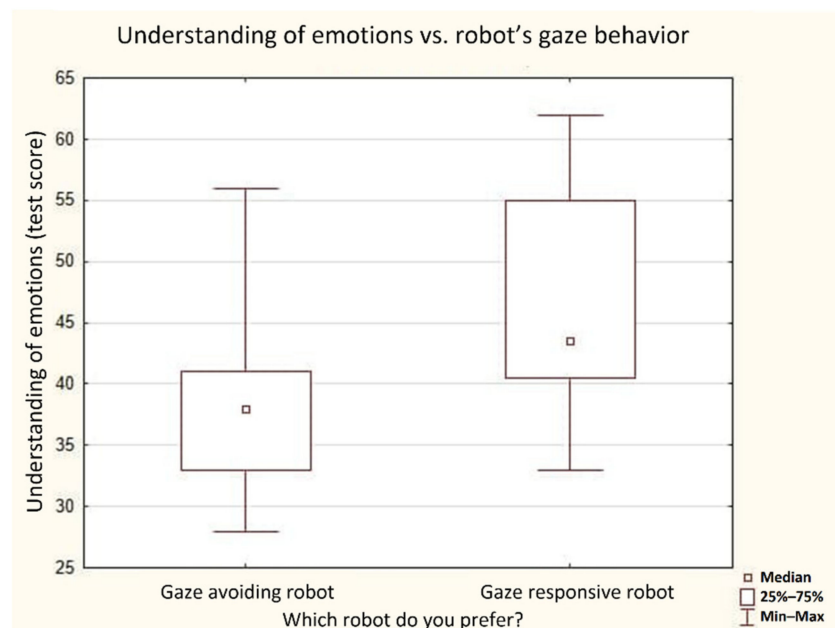


Figure 6. Level of emotional intelligence in our subjects vs. their preferences of the robots' social gaze behavior.

The results show that people with high sensitivity to emotions, measured with a dedicated scale of emotional intelligence, preferred the robot with a responsive gaze ($p < 0.05$, Mann–Whitney U-test). Several people in this group even described the gaze-avoiding robot as reacting to the user's gaze by looking sideways and characterized it as shy, thoughtful, or even female. Other respondents noted that they were eager to attract the attention of the gaze-avoiding robot, as they believed the robot was losing interest during the narrative. Several respondents noted that the triangle robot seemed to be simultaneously (a) paying attention to the human's narrative and (b) thinking about the events of the story or addressing some intrusive thoughts. For example, some people reported that "it is interesting how the robot seems to look at me while looking sideways". We can suggest the following interpretation of this phenomenon. The participants were perceiving the gaze-avoiding movement of the triangle robot as a compound pattern in which the robot (a) moves his gaze away immediately after the user's gaze, thus showing a type of responsive-gaze behavior, that is, paying attention to the user, and (b) looks sideways, thus showing a pattern typical for thinking about the events, according to the REC corpus. In other words, for these participants, the immediate start and the pattern of the movement (looking sideways) were invoked by two different internal states assigned to the robot. The immediate start of gaze movement was interpreted as expressed attention, and the looking sideways pattern was interpreted as thoughtfulness.

Although gaze-avoiding and gaze-responsive behaviors were designed as straightforward reactions to the user's gaze, several people noted different aspects of the movements. Some people who preferred the gaze-avoiding robot characterized it as more expressive and active since it was moving between three positions with high amplitude. People could also concentrate not on the responsive phase (when the robot looked up or sideways), but on the return movement to the inactive position (when the robot was looking down), which was sometimes interpreted as being attentive to the cards on the table or being upset about the events of the story.

Assuming these deviations in the descriptions, the evaluation of the group with the correct identification of the differences between the robots is rather speculative. As for the criteria, we selected people who explicitly indicated that the robot with a square was looking at me and/or the robot with a triangle was looking sideways. The rate of people who gave this response was 8% (1 of 12 persons) in the gaze-avoiding preference group, 23% (5 of 21 persons) in the neutral group, and 69% (9 of 13 persons) in the gaze-responsive preference group. It means that people who better recognized the difference between the gaze-responsive and gaze-aversive robots also significantly often ($p < 0.01$, Spearman correlation) preferred the robot with the gaze-responsive behavior. Thus, people with high sensitivity to the emotions of others better recognized the gaze behavioral types of the robots and preferred the robot with the responsive-gaze behavior.

4. General Discussion

Our results clearly evidence that one of the core features of the social gaze, when simulated by robots, is that it can provoke the human to assign cognitive and emotional properties to the robot. While this is not a new result in the field of human–robot interactions, some novel facts can be reported from our experiments. For example, although social gaze in robots makes their perception more anthropomorphic, this influence often remains only implicit. As a matter of fact, gaze can be perceived as a latent cue; people systematically react to it in special problem tasks, but half of them do not mention this feature in self-reports explicitly. Ordinary people without any noticeable neurological symptoms also demonstrated substantial individual differences. The responsive gaze was especially significant for people with a high level of emotional intelligence but not for all the users. This is the next novel result of our study. In particular, cases of aversive gazing can well be interpreted positively and increase mutual social rapport within an evolving joint activity, as in our experiments on problem solving and storytelling. Previous works have concentrated exclusively on momentary eye-to-eye contacts as the prime indication of the joint attention state ([5,22,44], among others). Therefore, researchers have ignored the necessary balancing of several behavioral acts by means of eye movements in a continuous interaction over time.

As perceived in its dynamics, gaze and the change of gaze direction can be described as a change of the object of interest (a more traditional approach) or as a change of the prevailing cognitive state, which controls the gaze. Thus, eye movements communicate at least two cognitive states—initial and final. We can consider this competition to be among the essential features of consciousness processes described by some authors as a competition and suppression of habitual strategies of behavior and cognitive states [45]. Not accidentally, people with high emotional intelligence can note and interpret minor features of the gaze (gaze response and gaze direction), assigning to it several cognitive states (“he heard me, but is thinking about something else”). The competitive gaze simulation can serve as a core method to communicate the conscious state of the agent by provoking listeners to attribute consciousness to the emotional robots during the communication.

Further perspectives in this line of research concern the imitation of phenomena such as the eye–voice span (eyes are ahead of the voice while reading aloud) and two main modes of eye movements and vision—one related to ambient vision (as in visual search) and one related to the focal mode of cognition and volition (as in identification and decision making). Although gaze was traditionally studied within the mechanism of attention of a

subject, the modern approach to gaze activity as a means of communication can suggest new concepts and allow for creating a new generation of natural emotional interfaces. In further developments, we strongly hope for improvement in the registration of eye movements and social gaze beyond the current methodology. This will open the way to their broader use in engineering as interfaces for automated systems for data processing and control, such as the gaze–brain–computer interface to control the internet of robotic things or eye–hand coordination to replace the overwhelming, but not always precise (and sometimes even impossible), computer mouse movements. One example of a case in which the gaze is more precise and faster than the computer mouse movements is the localization of objects from within a speedy vehicle or in conditions of ordinary industrial production.

As can be seen particularly from the results of Experiments 2 and 3, several behavioral patterns could be executed on the robot at the same time. This was especially the case if their BMLs referred to complementary effectors, when, for example, a responsive gaze was performed by the eyes and head, speech production controlled the mouth and a pointing gesture was performed by a hand. An external stimulus can also invoke several scripts, each of which can suggest their behavioral pattern for the execution. This combination of patterns allows developers to create compound and rich behavior in the robot and even simulate some emotional blending when the activated scripts represent contradictory emotions [46]. This type of blending corresponds to the architecture of reflexive effects of consciousness, simulated earlier with the help of the scripts mechanism, where the first, most relevant script, is accompanied by secondary scripts that are seemingly less relevant but provide alternative views on the situation perceived and to be managed [47].

We do not currently have direct empirical evidence for this idea. However, indirectly, it is supported by the bulk of neurophysiological, neurolinguistic, and even neuromolecular data, disputing the old concept that the left cerebral hemisphere in humans “dominates” over the “subdominant” right hemisphere [48,49]. At least in a clinical context, disorders of spatial, corporeal, emotional, and self-related consciousness primarily result from lesions of the right brain hemisphere [50,51]. A neurolinguistic pendant to these data is the well-established knowledge that while the left hemisphere supports explicit linguistic functions in most humans, the right prefrontal cortex is important for the implicit understanding of specifically human communication pragmatics, metaphorical language, humor, irony, and sarcasm, as well as eye-to-eye contact [52–54]. Moreover, dedicated brain “machinery” seems to be behind these cognitive–affective effects with a specific course of neurodevelopment in early ontogenesis [55]. The role of gaze contact is different for children with brain-and-mind disturbances, such as autism spectrum disorders and Williams syndrome [56,57].

However, the pathway to reliable comprehension and modeling of this neurophysiological phenomenology is difficult and conceivably long. In terms of needed usability, computational solutions that have been delineated in the present article are more practical and feasible.

5. Conclusions

Social gaze is a compound phenomenon, as it can be controlled by different cognitive states and, thus, convey information on the rich internal organization of the subject. As we have shown, robots that control their gaze via a computational model of competition between several internal states are perceived as more emotional and invoke higher empathy in humans. In the situation of storytelling by the robot, this competition arose between the attention paid to the opponent (human), the politeness strategy (to avoid long gazes), and simulated thoughtfulness, thus manifesting courtesy and reflection on the part of the robot. In the situation of puzzle-solving, this conflict arose between the simulated attention to the game elements for the next successful move and the opponent who should understand the robot’s instruction; this can manifest in the engagement of the robot to the process of problem-solving and be interpreted as its “intention” to cooperate with the human. In the situation of story listening, the attention to the user’s face and gaze manifested

the engagement of the robot to the story narrated by the user, while the preference of the gaze-responsive partner was correlated with the human ability to recognize others' emotions. In this respect, the social gaze system can be considered not only a system of visual perception attention and action but as a significant component to establishing fluent communication between robots and humans. Understandably, it is the key technology to make the users attribute human internal states to the robot in natural communication.

On a more conceptual level, one can consider essential features of consciousness as contrasted to the automated processes. Here, again, we used a minimalistic approach to try to find the minimal architecture of computation, demonstrating the essential features of reflexivity [47]. Interestingly, such architectures seem to be relatively simple, but they probably demand a kind of working memory extension to consider responses that seem to be not necessarily the most appropriate in the current context at the first sight. In our view, those are the "second-choice" resources for producing a robot's responses that will induce in humans an impression of being intelligent and even conscious. Though starting this endeavor with elementary hardware and processing architectures, we think that a computational solution to our fluent and partner interaction with artificial agents may be by far more feasible than in wet or in silico brain modeling.

Supplementary Materials: A video with the demonstration of the experiments is available online at www.youtube.com/channel/UCaJ7WUJb_NuqELyJyUrp45Q (accessed on 1 November 2021).

Author Contributions: Conceptualization, B.M.V. and A.K.; methodology, B.M.V., A.Z. and A.K.; software, N.A. and K.K.; formal analysis, A.K. and A.Z.; investigation, A.Z., L.Z. and A.K.; writing—original draft preparation, A.K., A.Z. and L.Z.; writing—review and editing, B.M.V. and A.K.; visualization, A.K.; supervision, B.M.V.; project administration, A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The first two experiments were in part supported by the National Research Center "Kurchatov Institute" (decision 1057 from 2 July 2020). The development of the general model (part 2) and Experiment 3 (Section 3.3 and corresponding results) were supported by the Russian Science Foundation, grant number 19-18-00547.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the local Ethics Committee of National Research Center "Kurchatov Institute" (Protocol No.10 from the 1 August 2018).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Experimental data is available online at <https://bit.ly/3gjqV8b> (accessed on 1 November 2021).

Acknowledgments: We appreciate the administrative and technical support from the Russian State University for the Humanities in conducting a series of human–robot experiments during the time of the pandemic.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Velichkovsky, B.M. Communicating Attention: Gaze Position Transfer in Cooperative Problem Solving. *Pragmat. Cognit.* **1995**, *3*, 199–223. [[CrossRef](#)]
2. Pagnotta, M.; Laland, K.N.; Coco, M.I. Attentional Coordination in Demonstrator-Observer Dyads Facilitates Learning and Predicts Performance in a Novel Manual Task. *Cognition* **2020**, *201*, 104314. [[CrossRef](#)]
3. Beyan, C.; Murino, V.; Venture, G.; Wykowska, A. Editorial: Computational Approaches for Human-Human and Human-Robot Social Interactions. *Front. Robot. AI* **2020**, *7*, 55. [[CrossRef](#)] [[PubMed](#)]
4. Li, M.; Guo, F.; Ren, Z.; Duffy, V.G. A visual and neural evaluation of the affective impression on humanoid robot appearances in free viewing. *Int. J. Ind. Ergon.* **2021**, 103159. [[CrossRef](#)]
5. Schrammel, F.; Pannasch, S.; Graupner, S.T.; Mojzisch, A.; Velichkovsky, B.M. Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. *Psychophysiology* **2009**, *46*, 922–931. [[CrossRef](#)] [[PubMed](#)]

6. Ito, J.; Yamane, Y.; Suzuki, M.; Maldonado, P.; Fujita, I.; Tamura, H.; Grün, S. Switch from ambient to focal processing mode explains the dynamics of free viewing eye movements. *Sci. Rep.* **2017**, *7*, 1082. [[CrossRef](#)] [[PubMed](#)]
7. Velichkovsky, B.M.; Korosteleva, A.N.; Pannasch, S.; Helmert, J.R.; Orlov, V.A.; Sharaev, M.G.; Velichkovsky, B.B.; Ushakov, V.L. Two Visual Systems and Their Eye Movements: A Fixation-Based Event-Related Experiment with Ultrafast fMRI Reconciles Competing Views. *STM* **2019**, *11*, 7–16. [[CrossRef](#)]
8. Brown, P.; Levinson, S.C. *Politeness: Some Universals in Language Usage (Studies in Interactional Sociolinguistics)*; Cambridge University Press: Cambridge, UK, 1987.
9. Ekman, P.; Friesen, W.V. *Facial Action Coding System: A Technique for the Measurement of Facial Movement*; Consulting Psychologists Press: Palo Alto, CA, USA, 1978; ISBN 0931835011.
10. Müller, C. Gesture and Sign: Cataclysmic Break or Dynamic Relations? *Front. Psychol.* **2018**, *9*, 1651. [[CrossRef](#)]
11. Iriskhanova, O.K.; Cienki, A. The Semiotics of Gestures in Cognitive Linguistics: Contribution and Challenges. *Vopr. Kogn. Lingvist.* **2018**, *4*, 25–36. [[CrossRef](#)]
12. Admoni, H.; Scassellati, B. Social Eye Gaze in Human-Robot Interaction: A Review. *J. Hum. Robot Interact.* **2017**, *6*, 25–63. [[CrossRef](#)]
13. Scassellati, B. Mechanisms of Shared Attention for a Humanoid Robot. *Embodied Cogn. Action: Pap. 1996 Fall Symp.* **1996**, *4*, 21.
14. Breazeal, C.; Scassellati, B. A Context-Dependent Attention System for a Social Robot. In Proceedings of the IJCAI International Joint Conference on Artificial Intelligence, San Francisco, CA, USA, 31 July–6 August 1999; Volume 2.
15. Kozima, H.; Ito, A. Towards Language Acquisition by an Attention-Sharing Robot. In Proceedings of the Joint Conferences on New Methods in Language Processing and Computational Natural Language Learning, Sydney, Australia, 11–17 January 1998; pp. 245–246.
16. Al Moubayed, S.; Skantze, G. Perception of Gaze Direction for Situated Interaction. In Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction, Gaze-In 2012, Santa Monica, CA, USA, 26 October 2012.
17. Normoyle, A.; Badler, J.B.; Fan, T.; Badler, N.I.; Cassol, V.J.; Musse, S.R. Evaluating Perceived Trust from Procedurally Animated Gaze. In Proceedings of the Proceedings-Motion in Games 2013, MIG 2013, Dublin, Ireland, 6–8 November 2013.
18. Lehmann, H.; Roncone, A.; Pattacini, U.; Metta, G. Physiologically Inspired Blinking Behavior for a Humanoid Robot. In *Lecture Notes in Computer Science*; Including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics; Springer: Cham, Switzerland, 2016; Volume 9979.
19. Yoshikawa, Y.; Shinozawa, K.; Ishiguro, H.; Hagita, N.; Miyamoto, T. The Effects of Responsive Eye Movement and Blinking Behavior in a Communication Robot. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006.
20. Yoshikawa, Y.; Shinozawa, K.; Ishiguro, H.; Hagita, N.; Miyamoto, T. Responsive Robot Gaze to Interaction Partner. In Proceedings of the Robotics: Science and Systems, Atlanta, GA, USA, 27–30 June 2007; Volume 2.
21. Huang, C.M.; Thomaz, A.L. Effects of Responding to, Initiating and Ensuring Joint Attention in Human-Robot Interaction. In Proceedings of the Proceedings-IEEE International Workshop on Robot and Human Interactive Communication, Atlanta, GA, USA, 31 July–3 August 2011.
22. Fedorova, A.A.; Shishkin, S.L.; Nuzhdin, Y.O.; Velichkovsky, B.M. Gaze Based Robot Control: The Communicative Approach. In Proceedings of the International IEEE/EMBS Conference on Neural Engineering NER, Montpellier, France, 22–24 April 2015; Volume 2015, pp. 751–754.
23. Minsky, M.L. *The Society of Mind*; Touchstone Book: New York, NY, USA, 1988.
24. Allen, S.R. Concern Processing in Autonomous Agents. Ph.D. Thesis, University of Birmingham, Birmingham, UK, 2001.
25. Kotov, A.; Zinina, A.; Filatov, A. Semantic Parser for Sentiment Analysis and the Emotional Computer Agents. In Proceedings of the AINL-ISMW FRUCT 2015, Saint Petersburg, Russia, 9–14 November 2015; pp. 167–170.
26. Baker, C.F.; Fillmore, C.J.; Lowe, J.B. *The Berkeley FrameNet Project*; Association for Computational Linguistics: Stroudsburg, PA, USA, 1998.
27. Lyashevskaya, O.; Kashkin, E. Framebank: A Database of Russian Lexical Constructions. In *Communications in Computer and Information Science*; Springer: Berlin/Heidelberg, Germany, 2015.
28. Fillmore, C.J. The Case for Case. In *Universals in Linguistic Theory*; Bach, E., Harms, R.T., Eds.; Holt, Rinehart & Winston: New York, NY, USA, 1968; pp. 1–68.
29. Wierzbicka, A. *Semantic Primitives*; Athenäum: Frankfurt, Germany, 1972; ISBN 376104822X.
30. Shvedova, N.Y. *Russian Semantic Dictionary. Explanatory Dictionary, Systematized by Classes of Words and Meanings*; Azbukovnik: Moscow, Russia, 1998. (in Russian)
31. Kotov, A.A.; Zaidelman, L.Y.; Arinkin, N.A.; Zinina, A.A.; Filatov, A.A. Frames Revisited: Automatic Extraction of Semantic Patterns from a Natural Text. *Comput. Linguist. Intelle. Technol.* **2018**, *17*, 357–367.
32. Vilhjálmsón, H.; Cantelmo, N.; Cassell, J.; Chafai, N.E.; Kipp, M.; Kopp, S.; Mancini, M.; Marsella, S.; Marshall, A.; Pelachaud, C.; et al. The Behavior Markup Language: Recent Developments and Challenges. In *Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 99–111.
33. Kopp, S.; Krenn, B.; Marsella, S.; Marshall, A.; Pelachaud, C.; Pirker, H.; Thórisson, K.; Vilhjálmsón, H. Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In *Intelligent Virtual Agents*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 205–217.

34. Kotov, A.A.; Budyanskaya, E. The Russian Emotional Corpus: Communication in Natural Emotional Situations. In *Computer Linguistics and Intellectual Technologies*; RSUH: Moscow, Russia, 2012; Volume 1, Issue 11 (18), pp. 296–306.
35. Zinina, A.; Zaidelman, L.; Arinkin, N.; Kotov, A.A. Non-Verbal Behavior of the Robot Companion: A Contribution to the Likeability. *Procedia Comput. Sci.* **2020**, *169*, 800–806. [[CrossRef](#)]
36. Tsfasman, M.M.; Arinkin, N.A.; Zaydelman, L.Y.; Zinina, A.A.; Kotov, A.A. Development of the oculomotor communication system of the F-2 robot based on the multimodal REC housing (in Russian). In Proceedings of the The Eighth International Conference on Cognitive Science: Abstracts of Reports, Svetlogorsk, Russia, 18–21 October 2018; pp. 1328–1330.
37. Zinina, A.; Arinkin, N.; Zaydelman, L.; Kotov, A.A. The Role of Oriented Gestures during Robot’s Communication to a Human. *Comput. Linguist. Intelect. Technol.* **2019**, *2019*, 800–808.
38. Kotov, A.; Zinina, A.; Arinkin, N.; Zaidelman, L. Experimental study of interaction between human and robot: Contribution of oriented gestures in communication. In Proceedings of the XVI European Congress of Psychology, Moscow, Russia, 2–5 July 2019; 2019; p. 1158.
39. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [[CrossRef](#)]
40. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2014**, 1867–1874. [[CrossRef](#)]
41. Terzakis, G.; Lourakis, M. A Consistently Fast and Globally Optimal Solution to the Perspective-n-Point Problem. In *Computer Vision—ECCV 2020. ECCV 2020. Lecture Notes in Computer Science*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer: Cham, Switzerland, 2020; Volume 12346. [[CrossRef](#)]
42. Lyusin, D.V. A New Technique for Measuring Emotional Intelligence: The EmIn Questionnaire (in Russian). *Psychol. Diagn.* **2006**, *4*, 3–22.
43. Lyusin, D.V. EMIN Emotional Intelligence Questionnaire: New psychometric data (in Russian). In *Social and Emotional Intelligence: From Models to Measurements*; Institute of Psychology of the Russian Academy of Sciences: Moscow, Russia, 2009; pp. 264–278.
44. Iwasaki, M.; Zhou, J.; Ikeda, M.; Koike, Y.; Onishi, Y.; Kawamura, T.; Nakanishi, H. “That Robot Stared Back at Me!”: Demonstrating Perceptual Ability Is Key to Successful Human–Robot Interactions. *Front. Robot. AI* **2019**, *6*, 85. [[CrossRef](#)]
45. Posner, M.I. (Ed.) *Cognitive Neuroscience of Attention*; The Guilford Press: New York, NY, USA, 2004.
46. Ochs, M.; Niewiadomski, R.; Pelachaud, C.; Sadek, D. Intelligent Expressions of Emotions. In *International Conference on Affective Computing and Intelligent Interaction*; Tao, J., Tan, T., Picard, R.W., Eds.; ACII 2005, LNCS 3784; Springer: Berlin/Heidelberg, Germany, 2005; pp. 707–714.
47. Kotov, A.A. A Computational Model of Consciousness for Artificial Emotional Agents. *Psychol. Russia: State Art* **2017**, *10*, 57–73. [[CrossRef](#)]
48. Velichkovsky, B.M.; Krotkova, O.A.; Kotov, A.A.; Orlov, V.A.; Verkhlyutov, V.M.; Ushakov, V.L.; Sharaev, M.G. Consciousness in a Multilevel Architecture: Evidence from the Right Side of the Brain. *Conscious. Cogn.* **2018**, *64*, 227–239. [[CrossRef](#)] [[PubMed](#)]
49. Velichkovsky, B.M.; Nedoluzhko, A.; Goldberg, E.; Efimova, O.; Sharko, F.; Rastorguev, S.; Krasivskaya, A.; Sharaev, M.; Korosteleva, A.; Ushakov, V. New Insights into the Human Brain’s Cognitive Organization: Views from the Top, from the Bottom, from the Left and, particularly, from the Right. *Procedia Comput. Sci.* **2020**, *169*, 547–557. [[CrossRef](#)]
50. Howard, I.P.; Templeton, W.B. *Human Spatial Orientation*; Wiley: London, UK, 1966.
51. Harrison, D.W. *Brain Asymmetry and Neural Systems: Foundations in Clinical Neuroscience and Neuropsychology*; Springer: Bern, Switzerland, 2015.
52. Shammi, P.; Stuss, D.T. Humour Appreciation: A Role of the Right Frontal Lobe. *Brain* **1999**, *122*, 657–666. [[CrossRef](#)] [[PubMed](#)]
53. Schilbach, L.; Helmert, J.R.; Mojzisch, A.; Pannasch, S.; Velichkovsky, B.M.; Voegeley, K. Neural Correlates, Visual Attention and Facial Expression during Social Interaction with Virtual Others. In Proceedings of the 27th Annual Conference of Cognitive Science Society, Stresa, Italy, 21–23 July 2005; Bara, B.G., Barsalou, L., Bucciarelli, M., Eds.; Lawrence Erlbaum: Mahwah, NJ, USA, 2005; pp. 74–86.
54. Kaplan, J.A.; Brownell, H.H.; Jacobs, J.R.; Gardner, H. The Effects of Right Hemisphere Damage on the Pragmatic Interpretation of Conversational Remarks. *Brain Lang.* **1990**, *38*, 315–333. [[CrossRef](#)]
55. Jones, W.; Klin, A. Attention to eyes is present but in decline in 2-6-month-old infants later diagnosed with autism. *Nature* **2013**, *504*, 427–431. [[CrossRef](#)] [[PubMed](#)]
56. Baron-Cohen, S. *Mindblindness: An Essay on Autism and Theory of Mind*; MIT Press: Cambridge, MA, USA, 1995.
57. Riby, D.M.; Hancock, P.J.; Jones, N.; Hanley, M. Spontaneous and cued gaze-following in autism and Williams syndrome. *J. Neurodev. Disord.* **2013**, *5*, 13. [[CrossRef](#)] [[PubMed](#)]