



# Social Eye Gaze in Human-Robot Interaction: A Review

Henny Admoni and Brian Scassellati

Department of Computer Science, Yale University

---

This article reviews the state of the art in social eye gaze for human-robot interaction (HRI). It establishes three categories of gaze research in HRI, defined by differences in goals and methods: a human-centered approach, which focuses on people's responses to gaze; a design-centered approach, which addresses the features of robot gaze behavior and appearance that improve interaction; and a technology-centered approach, which is concentrated on the computational tools for implementing social eye gaze in robots. This paper begins with background information about gaze research in HRI and ends with a set of open questions.

*Keywords:* eye gaze, nonverbal behavior, social robotics, teamwork, mental models, deixis, conversation, collaboration

---

## 1. Introduction

The field of Human-Robot Interaction (HRI) strives to enable easy, intuitive interactions between people and robots. Such interactions require natural communication. Although verbal communication tends to be primary in human-human interactions, nonverbal behaviors, such as eye gaze (Argyle, 1972) and gestures (McNeill, 1992), can convey mental state, augment verbal communication, and reinforce what is being said (Goldin-Meadow, 1999). Eye gaze is a particularly important nonverbal signal—compared with pointing, body posture, and other behaviors—because evidence from psychology suggests that eyes are a cognitively special stimulus, with unique “hard-wired” pathways in the brain dedicated to their interpretation (Emery, 2000).

The earliest research into communicative gaze was led by the virtual agent community in the 1990s (Cassell, Torres, & Prevost, 1998; Thórisson, 1994, for example). Virtual agents were imbued with eye gaze as a means of capturing attention, maintaining engagement, and increasing conversational fluidity with human users (Cassell, 2000). Roboticists began introducing meaningful eye gaze into their systems in the late 1990s, in robots such as Cog (Scassellati, 1996), Kismet (Breazeal & Scassellati, 1999b), and Infanoid (Kozima & Ito, 1998).

Modern-day approaches to incorporating eye gaze into human-robot interactions vary widely; research investigating the effects of social eye gaze on human-robot interactions spans the fields of robotics, virtual agents, artificial intelligence, and psychology. Some researchers use robots as stimuli to understand the limits of human perception. Others try to understand the effects of robot gaze

---

Authors retain copyright and grant the Journal of Human-Robot Interaction right of first publication with the work simultaneously licensed under a Creative Commons Attribution License that allows others to share the work with an acknowledgement of the work's authorship and initial publication in this journal.

by manipulating features of robot appearance and behavior and measuring their influence on human responses. Still others focus on the underlying technologies required for establishing convincing social eye gaze.

In this review, we present the current state of research on social eye gaze in human-robot interaction. To address the large variety of research included in this topic, we divide the corpus of work on gaze in HRI into three broad categories of research. The categories are distinguished both by their goals and by their methods. These categories are as follows:

**Human-focused:** This research aims to understand the characteristics of human behavior during interactions with robots. The focus is on the features and limits of human behavior and perception, with the robot serving as a stimulus to provoke a measurable response. This research generally involves well-controlled, laboratory-based studies.

**Design-focused:** This research investigates how design choices about a robot, such as its appearance or behavior, can impact interactions with humans. Design-focused papers tend to manipulate one feature of robot gaze behavior at a time (such as the length of fixation) to reveal people’s response to that feature and include both laboratory-based and field-based evaluations.

**Technology-focused:** This research aims to build computational tools for generating robot eye gaze in human-robot interactions. Though the technologies may be evaluated with human users, this work generally focuses on mathematical or technical contributions, rather than the effects of the system on the interaction.

These categories represent one way to segment the research around eye gaze in HRI, but they do not represent mutually exclusive areas. A single study may contribute in multiple categories, such as an evaluation of a data-driven model of conversation through a laboratory-based human study. In this review, we divide research literature by primary category, the area in which the paper’s contribution is primarily focused.

The focus of this review is *social eye gaze*, that is, any gaze that can be interpreted as communicative by an observer. Biological evidence suggests that the human eye has evolved to be especially capable of such social communication. Though other vertebrate species can recognize eye gaze and attention cues, humans have the unique ability, even beyond non-human primates, to infer others’ intentions from eye gaze (Emery, 2000). The unique morphology of the human eye—with a large, white sclera that clearly signals gaze position—enables this social signal (Kobayashi & Kohshima, 2001).

Social eye gaze includes eye movements that are intentionally expressive, such as gaze aversions that are designed to communicate thoughtfulness. Social eye gaze also includes eye movements that serve a purpose that is not *explicitly* communicative, such as orienting a robot’s field of view on an object of interest, as long as these movements are part of an interaction where they might be perceived by other people. Social eye gaze does not include eye movements that are not typically perceived by others during social interactions, such as gaze actions that happen in isolation, viewpoint-stabilization actions like the vestibulo-ocular reflex, or visual processing routines that do not involve changing the camera’s point of focus.

Throughout this review, we refer to various types of eye gaze using established terminology:

- *Mutual gaze* is often referred to colloquially as “eye contact”; it is eye gaze that is directed from one agent to another’s eyes or face, and vice versa. Face-directed gaze without reciprocity is not mutual gaze.

- *Referential gaze* or *deictic gaze* is gaze directed at an object or location in space. Such gaze sometimes occurs in conjunction with verbal references to an object, though it need not accompany speech.
- *Joint attention* involves sharing attentional focus on a common object (Moore & Dunham, 2014). It can have several phases, beginning with mutual gaze to establish attention, proceeding to referential gaze to draw attention to the object of interest, and cycling back to mutual gaze to ensure that the experience is shared.
- *Gaze aversions* are shifts of gaze away from the main direction of gaze, which is typically a partner's face. Gaze aversions can occur in any direction, though some evidence suggests the purpose of the aversion influences the direction of the shift (Andrist, Tan, Gleicher, & Mutlu, 2014).

The type of eye gazes a robot will use in a human-robot interaction will depend on the context and goals of the interaction. Eye gaze can reveal a social robot's mental states, including its knowledge and goals (Fong, Nourbakhsh, & Dautenhahn, 2003). Gaze can be used by socially assistive robots to demonstrate their engagement with and attention to a user (Tapus, Matarić, & Scassellati, 2007). Robot eye gaze can increase the fluidity of conversation (Mavridis, 2015) or direct a user's attention to relevant information in a tutoring setting (Johnson, Rickel, & Lester, 2000). However, a tutoring robot may want to express attention to and engagement with a user by performing frequent mutual gaze, while a collaborative assembly-line robot may prioritize task-focused gaze that enables joint attention and object reference.

The remainder of this review is organized around the three research categories established earlier: human-focused, design-focused, and technology-focused. First, Section 2 provides background about concepts and terminology that are common throughout the diverse studies described in this article. The review of current research begins in Section 3 with an introduction to gaze in human-human interactions, focusing on findings that are relevant to eye gaze for human-robot interactions. This section introduces insights from psychology that influence the development of gaze for robotics. Section 4 discusses human-focused research on gaze in HRI, including human capabilities and limitations when interacting with robots that use gaze communication. Section 5 describes design-focused research, specifically how a robot's physical appearance and behavior can be manipulated to elicit effective eye gaze communication within human-robot interactions. Section 6 presents technology-focused research, covering the various systems and frameworks for developing robot eye gaze. The paper concludes in Section 7 with questions for future research that will expand the understanding of eye gaze in HRI.

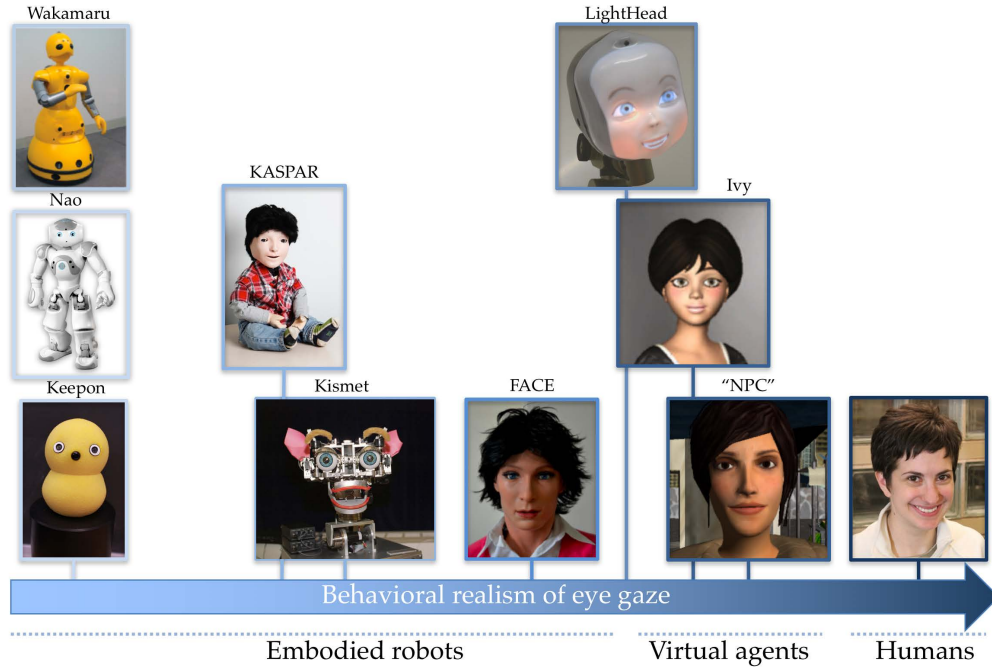
## 2. Background

This section describes some common themes found throughout the research on social eye gaze for HRI. In identifying the commonalities, this section also highlights the diversity in this body of work; many different approaches, domains, metrics, and technologies make up the state of the art in social eye gaze for HRI.

### 2.1 Robot appearance

Eye gaze research in HRI is conducted using robots with a wide range of variability in appearance and capability. These platforms range from simple cartoon-like robots to extremely lifelike humanoids and virtual agents.

The differences in gaze capabilities are related to the high cost of implementing eye movements in robots. Each movement along an axis, also known as a *degree of freedom*, must be produced by



*Figure 1.* Robots and virtual agents with a range of appearances and capabilities are used for gaze research in HRI. This spectrum roughly sketches the range of behavioral realism with examples drawn from research cited in this review: Wakamaru (Szafer & Mutlu, 2012), Nao (Aldebaran, 2015), Keepon (author photograph), KASPAR (courtesy of the Adaptive Systems Research Group, University of Hertfordshire, UK), Kismet (Breazeal & Scassellati, 1999a), FACE (Zaraki, Mazzei, Giuliani, & De Rossi, 2014), LightHead (Delaunay, 2015), Ivy (Andrist, Mutlu, & Gleicher, 2013), and an NPC (Normoyle et al., 2013).

some motor or other actuator. Adding capabilities means adding actuators, some of which must be quite small (to fit into the robot’s head) and powerful (to perform rapid movements like saccades). These requirements drive up a robot’s cost, complexity, and fragility. Most designers of social robots attempt to minimize these costs by choosing not to implement some biological capabilities.

Fig. 1 illustrates the spectrum of biologically realistic behaviors in robot eye gaze. This spectrum is a rough indicator of the range of human-likeness in eyes, in terms of behavioral capability. The extreme right end of the realism spectrum contains humans. Moving leftward on the spectrum indicates descending levels of behavioral realism, with fewer human-like capabilities such as pupil dilation, ocular torsion, and saccades.

Just below humans on the spectrum are virtual agents, which have the potential for extremely high levels of behavioral realism. By nature of being animated, virtual agents can mimic human eye capabilities with greater precision than physical robots, though computationally encoding biologically realistic gaze behavior is an active area of research (Ruhland et al., 2015). While some virtual agents are implemented with complex, biologically faithful models of muscle movement that control eye motion, others use motion generators that are less consistent with the underlying biology (Ruhland et al., 2015), so there is a range of possible realism within the virtual agent literature. In Fig. 1, the virtual agent referred to as “NPC” uses a biologically-based model to animate its saccades,

blinks, and gaze shifts (Normoyle et al., 2013). In contrast, the virtual agent called Ivy uses timings of gaze aversions drawn from video-coded observations of human conversation (Andrist, Mutlu, & Gleicher, 2013).

On the spectrum between virtual agents and embodied robots are retro-projected (or back-projected) robot faces. This modern technology projects an image onto the rear of a translucent molded mask. From the front, the image appears to be drawn directly onto the mask, providing the illusion of an embodied robot with greater animation flexibility. Due to the fixed shape of the mask, there are constraints on the dimensions of images that can be projected (Delaunay, Greeff, & Belpaeme, 2009), but researchers have implemented biologically-inspired gaze movements on retro-projected robot heads to elicit the perception of joint attention (Delaunay, Greeff, & Belpaeme, 2010) and gaze direction (Al Moubayed & Skantze, 2012).

Moving down the spectrum of behavioral realism, different capabilities are lost. Even the most realistic physical robots, for example, do not implement pupil dilation, though this behavior is an indicator of mental state (such as cognitive effort) in humans (Hyönä, Tammola, & Alaja, 1995). The behaviorally realistic robot pictured in Fig. 1, FACE, uses a human-like gaze model based on motion capture data from human examples to control the speed and magnitude of eye movements (Zaraki, Mazzei, Giuliani, & De Rossi, 2014).

Less behaviorally realistic robots retain gaze capabilities but have simpler appearances and gaze control models. Kismet has an independent pan and joint tilt degrees of freedom for each eye, two degrees of freedom for each eyebrow, and independent eyelids, enabling expressive behavior like winking (Breazeal, Hoffman, & Lockerd, 2004). Still less behaviorally realistic robots, such as KASPAR (Dautenhahn et al., 2009), have eyes that do not move independently of each other, eliminating the capability to perform lower-level components of biological gaze, such as vergence.

At the extreme low end of the realism spectrum are robots with fixed eyes. These robots, such as Keepon (Kozima, Michalowski, & Nakagawa, 2009), Nao (Aldebaran, 2015), and the Wakamaru robot (Szafir & Mutlu, 2012), are incapable of eye movements separate from head orientation, such as what people perform when orienting to a lateral visual target (Freedman & Sparks, 2000). Instead, these robots rely on head turns to indicate gaze direction. While this mechanism can be communicative on a gross level, there is evidence that head pose is an inadequate indicator of human gaze direction in human-robot interactions (Kennedy, Baxter, & Belpaeme, 2015b).

The variability in appearance and capability of robot eyes is important to note when discussing research on robot eye gaze. Because studies are conducted with different robots, their results may not directly transfer from one robot to another. Each study described in this review should be considered in the context of the robot or virtual agent it employs.

## 2.2 Embodiment and virtual agents

Much of the work on social eye gaze emerged from the virtual agents community in the 1990s (Cassell, Torres, & Prevost, 1998; Thórisson, 1994). This work led the way for embodied gaze research in robotics, and the virtual agent community continues to make advances in the understanding and design of social gaze for intelligent agents (Ruhland et al., 2015). For this reason, virtual agents are presented alongside physically embodied robot systems in this paper. However, there are some notable differences between the two fields.

Virtual agents can provide fine control over the appearance and timing of gaze behaviors, such as subtle eyelid, eyebrow, and eye ball movements. These types of fine movements are difficult to achieve with physical motors on embodied robots. Though some hyper-realistic humanoid robots—such as Geminoid (Sakamoto, Kanda, Ono, Ishiguro, & Hagita, 2007) and FACE (Zaraki, Mazzei, Giuliani, & De Rossi, 2014)—strive to achieve human-like face actuation, most do not achieve the level of facial expressiveness available in animated characters. Therefore, virtual agents provide a

platform with which to study the effects of well-controlled, subtly expressive motions of social eye gaze.

There is disagreement, however, on whether physically embodied systems are better for interactions than animated agents or even video representations of the physical systems. Some researchers have found that physically co-present embodied systems improve interactions over virtual systems (Li, 2015). Children spend more time looking at a robot tutor that is physically embodied than at a virtual representation of that robot (Kennedy, Baxter, & Belpaeme, 2015a), and adults retain lessons about a cognitive puzzle better when they had been tutored by a physically embodied robot than by a video representation of that robot (Leyzbeg, Spaulding, Toneva, & Scassellati, 2012). People also fulfill unusual requests from a robot more frequently when that robot is physically embodied than when it is telepresent (Bainbridge, Hart, Kim, & Scassellati, 2011), though the anthropomorphism of the embodiment may influence their willingness to do so (Bartneck, Bleeker, Bun, Fens, & Riet, 2010). Physically embodied agents are rated more positively (Powers, Kiesler, Fussell, & Torrey, 2007; Wainer, Feil-Seifer, Shell, & Matarić, 2007) and attributed greater social presence (Lee, Jung, Kim, & Kim, 2006) than their virtual or telepresent counterparts.

However, not all research has supported the benefit of physical embodiment over virtual presence. In a tutoring interaction involving sorting, children fail to show differences in learning from embodied and virtual robots (Kennedy, Baxter, & Belpaeme, 2015a). In an interaction with a health-care robot, people remembered less information provided by a physically co-located robot than information provided by a virtual representation of that robot (Powers, Kiesler, Fussell, & Torrey, 2007).

Research on embodiment to date has not specifically focused on the effect of embodied social gaze (see Section 7.3 for how this question might be addressed). Whether or not embodiment affects an interaction, research on both virtual agents and physically embodied robots is important for understanding social gaze for intelligent agents, and both the virtual agents and robotics communities have made important contributions to our understanding of eye gaze in human-agent interaction.

### 2.3 Study locations and controls

Human-robot interactions can be evaluated both inside and outside of the laboratory. Laboratory-based and field-based studies have complementary benefits and limitations, and both are important for investigating eye gaze in HRI. Based on the location of the study, researchers can control the environment and potential confounding variables to a greater or lesser degree. The trade-off for increased control is a decrease in the generalizability of the research findings to real-world settings.

Laboratories provide well-controlled environments in which to perform highly repeatable, consistent experiments. The laboratory can be outfitted with sensors to capture a variety of experimental data, including cameras for video (Bohus & Horvitz, 2010), skeleton tracking systems to detect body positions (Sorostinean, Ferland, Dang, & Tapus, 2014), and eye trackers for precise gaze analysis (Yu, Schermerhorn, & Scheutz, 2012). Laboratory-based studies are particularly well-suited to research that systematically manipulates a variable to understand its effect on an interaction, due to the capability of excluding potential confounding factors by rigidly controlling the environment.

Human-focused research is often performed in the laboratory, so that the conditions eliciting the measured human response are well-defined. For the same reason, many design-based studies are also conducted in a laboratory. However, laboratory-based studies are limited in their ecological validity, because the controlled and restricted environment does not necessarily represent how people and robots will operate in the real world. Thus, some design-based and technology-based studies choose to measure a robot's effect on interactions in the field.

Field-based studies involve placing robots in naturalistic environments, such as shopping malls (Satake et al., 2010), museums (Yamazaki, Yamazaki, Burdelski, Kuno, & Fukushima, 2010), and

building atriums (Knight & Simmons, 2013). Interactions tend to be more free-form, because the circumstances of the interactions cannot be precisely predicted or controlled. Data collection is often more limited than in laboratory-based studies and tends to be more observational than empirical. However, these types of studies can more accurately reveal people’s interactions with robots “in the wild.”

There is a spectrum of study types between these two extremes. For example, laboratories can be augmented with furniture to manufacture a more realistic setting (Pandey, Ali, & Alami, 2013). Naturalistic scenarios can be temporarily constructed with somewhat controlled conditions, for instance by evaluating a robot during a public demonstration (Bennewitz, Faber, Joho, Schreiber, & Behnke, 2005), where sensors can be arranged for additional data collection. In this paper, we include studies across this spectrum of study types, from carefully-controlled laboratory research to long-term deployments in everyday human environments (Simmons et al., 2011).

## 2.4 Evaluation metrics

When evaluating the effects of gaze on human-robot interactions, both objective and subjective metrics can provide useful information. Which evaluation metric is used depends on the interaction task and the research goals. This section provides an overview of the many objective and subjective measures used in research on gaze in HRI, with some specific examples of each.

*2.4.1 Objective Measures* Objective metrics often measure a user’s observable behavior. These metrics range in scale from millisecond-level measurements to broad observations of long-term behavior. High-level categories of objective metrics include measures of human behavior (e.g., eye movements) and performance (e.g., task completion time).

Precise measurements can reveal low-level (and not necessarily conscious) responses to robot gaze. For example, measuring millisecond-level response times to a robot’s directional gaze (Admoni, Bank, Tan, & Toneva, 2011) or recording tiny eye saccades with an eye tracker (Yu, Schermerhorn, & Scheutz, 2012) can reveal underlying differences between people’s responses to robots and humans.

Larger-scale measurements can quantify a robot’s effect on longer-term human behavior. For example, how well a robot’s referential gaze facilitates understanding of object references can be measured by how long it takes a user to select the correct object (Admoni, Datsikas, & Scassellati, 2014; Boucher et al., 2012; Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005). The effectiveness of a robot tutor’s gaze behaviors can be revealed by the amount of information a user is able to recall from the interaction (Andrist, Pejsa, Mutlu, & Gleicher, 2012a; Szafr & Mutlu, 2012). Information recall can also act as a proxy for attention: if participants pay more attention, they can recall more information, so measuring recall reveals how much attention different robot gaze behaviors elicit from people (Huang & Mutlu, 2012; Mutlu, Forlizzi, & Hodgins, 2006).

Some objective measures involve post-hoc interpretation of human behavior, often accomplished through video coding. This process entails careful analysis of a recorded interaction to evaluate users’ responses to a robot’s gaze behaviors, in terms pre-defined items like engagement behaviors (Karreman, Ludden, Dijk, & Evers, 2015), the conversational function of utterances (Andrist, Mutlu, & Gleicher, 2013), and use of body language (Huang & Mutlu, 2014). Because these post-hoc interpretations may be subject to the coder’s perceptions and biases, these interpretations are often coded by two or more individuals, with correlations confirmed by statistics, such as Cohen’s  $\kappa$ -coefficient (Cohen, 1960).

Objective evaluations can be applied to the robot systems themselves. For example, the success of a robot gaze system can be measured by whether a robot can predict the correct speaker (Trafton & Bugajska, 2008; Vertegaal, Slagter, Veer, & Nijholt, 2001) or influence human users into certain

conversational roles (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009).

**2.4.2 Subjective Measures** Subjective measures can provide insight into user experiences that may not be outwardly observable. Subjective measurements typically involve collecting user perceptions and opinions through surveys and interviews.

The most common type of subjective measure for studies investigating social eye gaze in HRI is a survey or questionnaire, often provided to users at the end of an experiment (Andrist, Mutlu, & Gleicher, 2013; Choi, Kim, & Kwak, 2013; Huang & Mutlu, 2014; Sidner, Kidd, Lee, & Lesh, 2004; Trafton & Bugajska, 2008, among many others). Survey questions are often formulated as Likert scales, through which participants reveal their perceptions and opinions by indicating their strength of agreement or disagreement with selected statements. For example, to evaluate how well gaze behaviors make a robot seem like a positive interaction partner, these scales measure characteristics like intelligence, animacy, and likeability (Bartneck, Kulić, Croft, & Zoghbi, 2009). Subjective measures can also include direct evaluations of a robot’s behavior. For example, to evaluate how well a robot can express emotions by changing its eye and facial expressions, a user might be asked to identify what emotion the robot is conveying for various expressions (Li & Mao, 2012b).

Interviews are another tool for eliciting subjective feedback from users. Interviews can reveal, for example, children’s subjective impressions of a robot tutor (Saerbeck, Schut, Bartneck, & Janse, 2010). Interviews can also be used to elicit anecdotal evidence that supports or explains the study’s findings (Huang & Mutlu, 2016; Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009).

Manipulation checks are a particular kind of measure that identifies whether an experimental manipulation was effective or not. In HRI, a manipulation check often ascertains whether participants consciously experienced the manipulation, which may be important for evaluating the validity of results. It can be given as a single item on a questionnaire (Huang & Mutlu, 2016; Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009) or as part of an interview (Admoni, Dragan, Srinivasa, & Scassellati, 2014; Zheng, Moon, Croft, & Meng, 2015). For example, in a study investigating how the duration of a robot’s gaze toward people affects their participation in a conversation, participants were explicitly asked how much the robot gazed at them and at their partner as a way of judging whether they actually perceived different durations of robot gaze (Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009).

Objective and subjective measures provide complementary approaches for evaluating the effects of robot gaze in human-robot interactions. The field of HRI uses a diverse set of measures, and understanding the role of these different types of metrics is important for interpreting the research in the field.

### 3. Gaze in Human-Human Interactions

Gaze is important to human-human interactions, because it is closely tied to what people are thinking and doing (Kleinke, 1986). People use their observations of others’ eye gaze to guide everything from conversation (Kleinke, 1986) to speech (Argyle & Cook, 1976) and attention (Frischen, Bayliss, & Tipper, 2007). In this section, we draw out specific research findings from psychology that have a direct impact on the design of social eye gaze for human robot interaction. The described findings are then applied to research appearing in later sections of this review. The studies in this section are aligned into three general topics:

- How people use eye gaze for conversation and speech (relevant to Sections 5.1 and 5.2)
- How people use eye gaze when they refer to and manipulate objects (relevant to Section 5.3 and 5.4)
- The psychophysics of eye gaze and how to measure gaze effects (relevant to Section 4.2)



### 3.1 Gaze for conversation and speech

Most of the early research on eye gaze has focused on the role of gaze in conversation (Argyle, 1972; Argyle & Cook, 1976; Argyle & Ingham, 1972; Cook, 1977; Kendon, 1967; Kleinke, 1986). During conversations, eye gaze can be used to convey information, regulate social intimacy, manage turn-taking, and convey social or emotional states (Kleinke, 1986).

People generally look at what they are attending to, and so gaze in conversation predicts the target of conversational attention (Cook, 1977). When someone is listening, the person they are looking at is likely the person being listened to (88% of the time) (Vertegaal, Slagter, Veer, & Nijholt, 2001). Similarly, when someone is speaking, they are often looking at the target of their speech (77% of the time) (Vertegaal, Slagter, Veer, & Nijholt, 2001), though listener-directed gaze can occur significantly less frequently than speaker-directed gaze (Cook, 1977). In general, gaze is directed at conversational partners approximately 60–80% of the time (Argyle & Ingham, 1972; Cappella & Pelachaud, 2002; Vertegaal, Slagter, Veer, & Nijholt, 2001).

Eye gaze signals when speakers want to maintain or relinquish the floor, indicates cognitive effort, and balances attention with intimacy (Andrist, Tan, Gleicher, & Mutlu, 2014; Kendon, 1967; Kleinke, 1986). Early researchers observed that during an exchange of speaking turn, the speaker usually looks at the listener at the end of their utterance and the listener generally looks away as they begin their speaking turn (Kendon, 1967). More modern researchers have been able to use technology such as video cameras to extract very specific timings for the gaze cues that are part of conversation (Andrist, Mutlu, & Gleicher, 2013; Oertel, Włodarczak, Edlund, Wagner, & Gustafson, 2012). They have found, for example, that speakers establish mutual gaze approximately 2.4 seconds before relinquishing the floor (Andrist, Mutlu, & Gleicher, 2013). Intimacy modulating gaze aversions tend to be short (between 1 and 2 seconds), while gaze aversions that signal cognitive effort (such as looking away while beginning a response to a question) are longer, at about 3.5 seconds (Andrist, Mutlu, & Gleicher, 2013).

Gaze between partners is dynamic and follows the task at hand. For example, the “reference-action sequence”—in which an instructor refers to an object and then a worker acts on that object—can be divided into five cyclically repeating phases, each with their own distinct gaze behaviors: pre-reference, reference, post-reference, action, and post-action (Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015). A worker’s gaze tends to follow the instructor’s gaze in the early and late phases, while the instructor’s gaze tends to follow the worker’s behaviors during the middle phases (post-reference and action) as the worker performs the task (Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015).

In addition to managing interpersonal interactions, gaze also relates directly to speech syntax. People often look away from their partner when beginning the *theme* of the sentence (which indicates what the sentence is about) and look toward their partner when beginning the *rheme* of the sentence (which provides information or exposition about the theme) (Cassell, Torres, & Prevost, 1998).

The conversation topic also influences gaze. When guiding a tour, people look between the exhibit and their audience, among other nonverbal behaviors, and these behaviors elicit engagement responses from the audience (Yamazaki, Yamazaki, Burdelski, Kuno, & Fukushima, 2010). People show less mutual gaze when their conversation involves high levels of intimate self-disclosure (Kang, Gratch, & Sidner, 2012). Two partners’ nonverbal behaviors, including their eye gazes, can be used to extract the context of an utterance during an interaction, such as conveying a fact or answering a question (Admoni & Scassellati, 2014) (Fig. 2).

Gaze durations during conversation are also affected by people’s personalities. Extroverts spend more time looking at their partner than introverts (Andrist, Mutlu, & Tapus, 2015). People are also more likely to speak when their conversational partner looks at them more often (Vertegaal & Ding, 2002). However, the amount of mutual gaze between conversational partners depends on



*Figure 2.* The gaze dynamics of student (left) and teacher can reveal the context of an interaction. Here, the teacher is conveying a spatial reference, performing face-directed gaze at the student while the student observes the deictic gesture (Admoni & Scassellati, 2014).

the interpersonal dynamics between the partners, not only on each partner's individual traits (Broz, Lehmann, Nehaniv, & Dautenhahn, 2012; Cappella & Pelachaud, 2002).

### 3.2 Gaze for object reference and manipulation

Eye gaze plays an important role in communicating about the environment. When people refer to objects around them, they look at those objects ahead of naming or manipulating them (Griffin & Bock, 2000; Hanna & Brennan, 2007; Yu, Schermerhorn, & Scheutz, 2012). People are very good at identifying the target of their partner's referential gaze and use this information to predict what their partner is going to say (Boucher et al., 2012; Staudte & Crocker, 2011).

When referring to objects or locations around them, people's gaze is closely tied to the content of their speech (Hanna & Brennan, 2007). Objects are typically fixated one second or less before they are named (Griffin & Bock, 2000; Yu, Schermerhorn, & Scheutz, 2012), though this may be slightly longer when speakers must search for the object (Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015).

Teams of people use eye gaze as a subtle, non-intrusive channel of communication (Shah & Breazeal, 2010). When partners refer to objects or locations in the environment, people use their partner's eye gaze to predict their partner's next verbal object reference and can more quickly respond to that reference (Boucher et al., 2012). In contrast, when access to a partner's eye gaze is restricted, people are slower at responding to their partner's referential communication (Boucher et al., 2012). If there is ambiguity in the object reference, gaze is a strong and flexible cue for eliminating uncertainty about referential expressions (Hanna & Brennan, 2007; Staudte & Crocker, 2011). Infants as young as 14 months show greater interest to objects that an adult looks at than objects the adult cannot see (Brooks & Meltzoff, 2002), suggesting that gaze sensitivity is developed early.

When people are manipulating objects, their eye gaze is similarly tied to their task and intended action. Eye gaze typically reaches the object of interest before any movement of the hands has started (Land & Hayhoe, 2001). Though people fixate the same object while they act on it, eyes often shift to the next object in the task sequence before the action is completed on the current object (Land & Hayhoe, 2001). These shifts of gaze to a new object often correspond to the start of a significant kinematic event on the current object; for example, gaze directed at an object to be grasped will shift away from that object just as the hand closes around it (Johansson, Westling,

Bäckström, & Flanagan, 2001). Objects not related to the task at hand are rarely fixated (Hayhoe & Ballard, 2005).

Gaze is also used to signal availability for interaction. When they must pass objects back and forth, object handovers between people rely on a receiver signaling readiness to receive an object by gazing at their partner (Strabala, Lee, Dragan, Forlizzi, & Srinivasa, 2012; Strabala et al., 2013). Interestingly, the giver in this interaction is not required to return the receiver's mutual gaze in order for the handover to occur successfully. Caregivers in a nursing home demonstrate their availability to their patients through broadly distributed gaze, and people naturally wait for caregivers to establish mutual gaze before requesting assistance (Yamazaki et al., 2007).

### 3.3 The psychophysics of eye gaze

Psychophysics—the branch of psychology that explores how physical stimuli influence mental processes—has explored the effects of eye gaze on a cognitive level. A well-developed series of studies suggest that directional eye gaze is processed differently in the brain than other directional stimuli, indicating that eyes may be a unique type of stimulus (Downing, Dodds, & Bray, 2004; Driver et al., 1999; Friesen, Ristic, & Kingstone, 2004; Kingstone, Tipper, Ristic, & Ngan, 2004; Ristic et al., 2005).

People are very highly tuned to others' gaze direction. Three-month-olds already shift their attention in the direction of an adult's gaze (Hood, Willen, & Driver, 1998). In adults, seeing someone's eyes directed laterally—even in a photograph—evokes rapid, reflexive attention shifts in the direction of the gaze (Hood, Willen, & Driver, 1998; Langton & Bruce, 1999).

A series of psychophysics experiments has tested this reflexive attention shift and found that it is resistant to conscious control. In these experiments, participants are shown a picture of a face gazing to one side. Even when are told that they should look in the opposite direction of the gaze, their attention is still drawn to the direction of the gaze in the first 500 milliseconds, an effect called *counterpredictive cueing* (Downing, Dodds, & Bray, 2004; Driver et al., 1999; Friesen, Ristic, & Kingstone, 2004).

Counterpredictive cueing experiments provide evidence that faces are special stimuli that are processed in unique cognitive pathways, because the counterpredictive cueing effect is not seen in response to non-gaze directional cues such as arrows (Friesen, Ristic, & Kingstone, 2004) or tongues (Downing, Dodds, & Bray, 2004). A single image can even be manipulated to elicit the counterpredictive cueing effect or not, depending on whether the image is presented as a face or a car (Ristic et al., 2005). Functional MRI studies show that a single image activates different brain pathways depending on whether it is presented as eyes or as a non-social directional image (Kingstone, Tipper, Ristic, & Ngan, 2004), further strengthening the idea that eyes are processed differently than other cues.

This effect might be explained in part by people's strong tendency to have a theory of mind for another person, that is, a belief that the person has knowledge, goals, and intentions of their own. Functional MRI studies reveal a significant overlap in the brain areas that process theory of mind and those that process directional eye gaze (Calder et al., 2002). In fact, observing someone signaling the presence of an object with referential gaze elicits the same neural response as observing someone physically reaching to grasp that object (Pierno et al., 2006), indicating that people use gaze as a powerful indicator of others' intentions.

## 4. Human-Focused Research

The studies described in this section focus on learning about the characteristics and limits of human perception through human interactions with robots. These studies generally take place in well-controlled laboratory environments, where the limitations and features of human perception can be

closely examined. Understanding how people perceive and respond to robot gaze—including what is effective and what is not—is the first step in developing gaze behaviors for robots.

#### 4.1 Human response to robot social gaze

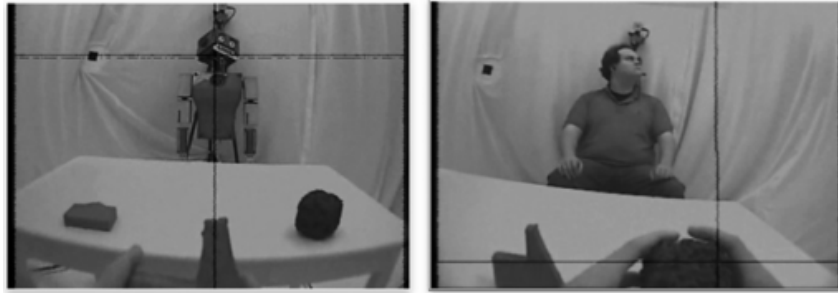
Before people can make use of a robot's social eye gaze, they must first perceive it. In multi-party conversations, people notice a robot's gaze when it looks at or near them, but not when it gazes at someone else nearby (Imai, Kanda, Ono, Ishiguro, & Mase, 2002). This suggests that the perception of robot gaze is egocentric—gaze is most frequently perceived when the robot is gazing directly at the viewer and is less frequently perceived when the robot is gazing at someone else. People have stronger feelings of “being looked at” when a robot gazes at them using short, frequent glances rather than longer, less frequent stares (Admoni, Hayes, Feil-Seifer, Ullman, & Scassellati, 2013).

People are also sensitive to robot eye gaze when that gaze is directed at objects or locations in the environment. For example, in object selection games, people can use referential gaze cues from a virtual agent (Bailly, Raidt, & Elisei, 2010) or a robot (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009) to make predictions about which objects to select, even when they are not consciously aware of those cues. Even ten-month old infants show sensitivity to robot gaze direction, following the line of sight of a non-humanoid robot head that demonstrates environmentally contingent behavior (Movellan & Watson, 2002). For back-projected robot heads, people can generally predict the target location of the robot's gaze, but their performance is worse with the robot head than with human eye gaze (Al Moubayed & Skantze, 2012; Delaunay, Greeff, & Belpaeme, 2010). Accuracy with the robot head suffers when the head is viewed from the side (Al Moubayed & Skantze, 2012; Delaunay, Greeff, & Belpaeme, 2010), when gaze involves just head orientation and not eye movement (Al Moubayed & Skantze, 2012), and when the projection is on a 3D sphere instead of a face-like mask shape (Delaunay, Greeff, & Belpaeme, 2010).

Such object-directed referential gaze has specific gaze timings that appear natural to people. Using an immersive virtual environment, researchers were able to empirically measure the timing of referential gaze during an interaction between a person and a virtual agent (Pfeiffer-Lessmann, Pfeiffer, & Wachsmuth, 2012). They found that the mean time a referential gaze dwelled on a referenced object was about 1.9 seconds, and that participants expected a responding gaze to be directed from their partner to the target object within about 2.5 seconds of their reference. These timing values can inform the production of gaze in future agent systems.

While people can successfully interpret robot eye gaze for object references, having a robot display mutual gaze also improves people's subjective and social evaluations of that robot. Mutual gaze from a stuffed animal companion robot leads to favorable evaluations of the robot (Yonezawa, Yamazoe, Utsumi, & Abe, 2007). When a robot is learning from human demonstration, displaying mutual gaze leads people to view the robot as more intentional than displaying random gaze; people spend more time teaching the robot, pay more attention to it, and speak more with it (Ito, Hayakawa, & Terada, 2004). However, social gaze produced by a robot is not necessarily required for people to attribute object ownership to that robot (Kanngiesser et al., 2015). In a cross-cultural study of children from China, Japan, and the UK, having a robot display independent motion and having a human perform social gaze toward the robot was enough for attributions of ownership to the robot, even absent any robot gaze cues.

People's preconceived expectations for an agent's gaze influence how they respond to that gaze. In what they describe as a “non-verbal Turing test,” researchers manipulated the amount of gaze following displayed by a virtual agent and asked participants to evaluate whether the agent was being controlled by a human partner or by a computer program (Pfeiffer, Timmermans, Bente, Vogeley, & Schilbach, 2011). They found that ascriptions of humanness varied by whether the human partner was introduced as naïve to the task, as cooperative, or as competitive, suggesting that interpretations



*Figure 3.* Comparing robot gaze to human gaze using eye tracking can reveal differences between people’s responses to the two stimuli. People spend more time looking at a robot partner’s face than a human partner’s face when naming an object (Yu, Schermerhorn, & Scheutz, 2012).

of gaze behavior “humanness” depend on the intent ascribed to the agent.

For robots that act as therapy assistants to children with autism spectrum disorder (ASD), gaze can be a particularly important cue because of the social gaze deficit that is often part of this disorder (Scassellati, Admoni, & Matarić, 2012). Some children with ASD show spontaneous social gaze behaviors in response to robots, including increased eye gaze and shared attention during robot interactions as compared to human interactions (Tapus et al., 2012). These findings lend support to the use of robots as therapy tools. However, there is large variability in responses, and other children do not demonstrate the same increase in gaze behavior (Tapus et al., 2012). Because gaze trajectories sometimes differ between people with ASD and those without, computationally recognizing and modeling people’s gaze might be a way to diagnose and evaluate ASD (Chawarska & Shic, 2009; Scassellati, 2007; Shic, Scassellati, Lin, & Chawarska, 2007).

#### 4.2 Differences in human response to robot and human gaze

It is tempting to assume that perfectly matching robot gaze behaviors to human gaze behaviors will elicit identical responses from people, but this is not always the case. Several studies suggest that gaze from robots is interpreted differently than gaze from humans.

In general, it is difficult to compare robot gaze to human gaze directly, because while robot gaze can be infinitely controlled, human gaze tends to have small, unpredictable variations. Several studies throughout this section directly compared robot gaze to human gaze through careful laboratory-based experiments. For example, one study made this comparison using a trained actor who performed identical behaviors to a pre-programmed robot (Fig. 3). While viewers’ gaze patterns were overall similar between the human and robot conditions, fine-grained analysis reveals differences in people’s responses to robot gaze and human gaze. For instance, people spend significantly more time looking at a robot partner’s face than at a human partner’s face when naming an object, indicating an apparent concern for ensuring that the robot is attending to the object in question (Yu, Schermerhorn, & Scheutz, 2012).

Other fine-grained analysis reveals that robot gaze is not afforded the same special cognitive status as human gaze. Recall from Section 3.3 that people show a tendency to unavoidably shift their attention in the direction of another person’s averted eye gaze, referred to as the reflexive cueing effect. This effect suggests that gaze is processed in a different neural pathway than other directional symbols like arrows. A test of reflexive cueing using both highly anthropomorphic and highly stylized robots showed that robots failed to elicit reflexive cueing in people, suggesting that robots

are cognitively processed more like arrows than like faces (Admoni, Bank, Tan, & Toneva, 2011; Admoni & Scassellati, 2012). In 12-month-old infants, eyetracking reveals anticipatory eye gaze shifts in response to human referential gaze but not to robot referential gaze (Okumura, Kanakogi, Kanda, Ishiguro, & Itakura, 2013a).

For infants, the interpretation of robot gaze seems to depend on whether that robot appears to be a social agent. When infants are shown videos of robots and humans looking at objects, they can follow the robot gaze as well as the human gaze. However, the infants look longer at, and show enhanced processing of, objects gazed at by the human but not objects gazed at by the robot (Okumura, Kanakogi, Kanda, Ishiguro, & Itakura, 2013b). Only after infants observe a robot engage in a socially communicative exchange with an adult do they follow a robot’s directional gaze (Meltzoff, Brooks, Shon, & Rao, 2010). Similarly, infants only imitate a robot’s intended action when it first establishes mutual gaze with an adult (Itakura et al., 2008). This suggests that, even to infants, robot gaze is not automatically as meaningful as human gaze, but it can be afforded meaningful status through social engagement.

The examples in this section provide some evidence for differences in human responses to robot versus human gaze, but more investigation is warranted (see Section 7.2). For example, while Yu *et al.*’s (2012) study tried to carefully control the human actor’s behavior to make it identical to the robot’s behavior, there may still have been minute differences in the human performance. Additionally, differences in appearance between the robot and human actor might have played a role in eliciting different gaze responses from people. Admoni *et al.*’s (2011) study employed two particular robots with specific appearances, and the characteristics of these appearances may have affected why robots did not elicit reflexive cueing. In the studies with infants, the robots exhibited few animacy cues in the non-social condition, which might have influenced whether infants saw them as intelligent agents. Possibly, some cue tangential to social communication, and not the eye gaze itself, led infants to respond differently to robots in these studies.

#### 4.3 Human-focused research summary

Studies focused on human perception of robot gaze have established that people can successfully identify the target of a robot’s gaze, whether it is looking at them or at other objects in the world. Human-focused studies have attempted to tease apart specific timings and gaze patterns that make robot gaze most effective, though more research in this area could establish specific timings or patterns of gaze that convey attention and object references most effectively.

Though simply orienting a robot’s gaze toward a particular location seems to be generally effective in conveying the intended target, perceptions of the robot’s animacy appear to modulate gaze effectiveness. People respond more favorably to robots that display socially contingent gaze, for example by establishing mutual gaze with their partners. Infants’ interpretation of robot gaze depends on whether they see the robot established as a social agent. And at very low levels of analysis (e.g., millisecond-level saccades), there seem to be different patterns of behavior toward gaze from robots and gaze from other humans. Additional research can further explore this magnifying effect of animacy on gaze (Section 7.2).

### 5. Design-Focused Research

Researchers who take the design-focused approach develop the appearances and behaviors of robots to address certain goals, such as demonstrating engagement or participating in joint attention. In this section, we describe how manipulations of robot gaze behavior can affect human-robot interaction both positively and negatively.

Gaze can serve many purposes, and the goal of eye gaze communication is often dictated by the task at hand. For example, a robot engaged in conversation might display user-focused mutual gaze,

while a tour guide robot performing a presentation might want to direct gaze to an exhibit using referential gaze. For this reason, we group the articles in this section by the task, or context, of the interaction: conversation, narration, collaboration, manipulation, and expression.

### 5.1 Conversation

Conversation involves an alternation of speaking and listening. For example, robots for tutoring or entertainment must be able to maintain an engaging, natural conversational exchange with human partners. The main challenges of conversation are managing attention and turn-taking between partners, selecting the correct gaze for the conversational content, and adopting the correct conversational roles.

Before beginning an interaction, a robot needs to gain the attention of its listeners. If the robot fails to successfully engage its intended partner, the listener can be unaware or uncertain about the robot's intent to communicate, even though they may be interested in that communication. Robots can use mutual gaze to improve the success of initiating conversation (Satake et al., 2010). Even a very simple illusion of gaze improves user attention. Simply having a virtual face on the flat-screen monitor of an embodied but non-anthropomorphic robot increases the number of users who stop when greeted by the robot (Bruce, 2002). Having the robot's head "look at" a person by turning toward the person's location has a similar effect on engagement, even without a virtual face, though the combination of face and tracking lead to the greatest user engagement (Bruce, 2002). Robots can also use gaze to acknowledge new arrivals to the conversation (Simmons et al., 2011).

Once the conversation has begun, conversational fluidity is managed as much by the absence of mutual gaze as by its presence. Gaze aversions can be used to demonstrate cognitive effort, modulate intimacy, and mediate turn taking (Andrist, Mutlu, & Gleicher, 2013). Using empirical timings for gaze aversions collected from lab-based observations of human-human conversation, researchers designed gaze aversion behaviors for virtual conversational partners. Virtual agents using gaze aversions for these conversational functions are more successful at regulating the conversational flow and elicit greater disclosure from people than agents that do not perform gaze aversions or perform gaze aversions at inappropriate times (Andrist, Mutlu, & Gleicher, 2013).

This gaze aversion model, when applied to embodied robots, yields a similar effect even though the robot (a Nao) uses head turns to signal gaze direction instead of articulated eyes (Andrist, Tan, Gleicher, & Mutlu, 2014). In some cases, though, averting gaze may not be the most effective way of mediating turn-taking. In an interaction that involved handing off speaking turns between a person and a Nao, flashing the eye LEDs to yield the speaking turn led to the fastest responses from people, while using gaze aversions actually led to slower responses than using no turn-taking cue at all (Van Schendel & Cuijpers, 2015).

People are also sensitive to the dynamic interplay between their own gaze and a robot's gaze. Robot gaze that is responsive to the user—that is, joint attention and mutual gaze that occur in response to human behavior—increases the self-reported "feeling of being looked at" over gaze that is independent of a user's behavior (Yoshikawa & Shinozawa, 2006; Yoshikawa, Shinozawa, Ishiguro, Hagita, & Miyamoto, 2006).

The content of conversation influences what kind of gaze works best. In conversations about emotionally neutral topics, robots that make eye contact are seen as more sociable and intelligent than robots that avoid it, but this effect is reversed when the topic of conversation is embarrassing, with eye contact avoiding robots rated more highly (Choi, Kim, & Kwak, 2013). In persuasive conversation, natural gaze behaviors improve a robot's persuasiveness (Ham, Cuijpers, & Cabibihan, 2015), even more than using expressive vocalizations (Chidambaram, Chiang, & Mutlu, 2012). Gaze also seems to mitigate the effects of other nonverbal behaviors on persuasiveness: When performed with eye gaze, persuasive gestures improve a robot's overall persuasiveness, but when per-

formed without eye gaze, persuasive gestures actually have the opposite effect, hindering a robot's persuasiveness (Ham, Cuijpers, & Cabibihan, 2015).

In multi-party conversations, robot eye gaze can influence how and when the robot or other people take on conversational roles. Several studies have found that a robot can use gaze behaviors to manipulate certain members of a group into taking conversational roles such as onlooker, active participant, or listener (Kirchner, Alempijevic, & Dissanayake, 2011; Mutlu, Kanda, Forlizzi, Hodgins, & Ishiguro, 2012; Mutlu, Shiwa, Kanda, Ishiguro, & Hagita, 2009). A robot can even use its gaze to take on those roles itself (Matsusaka, Fujie, & Kobayashi, 2001; Matsusaka et al., 1999). Robot gaze behaviors are successful at influencing people to conform to the intended roles as much as 97% of the time (Mutlu, Kanda, Forlizzi, Hodgins, & Ishiguro, 2012). A virtual agent's gaze can also influence which participant in a multi-party conversation takes the conversational floor next, with up to 86% effectiveness in releasing the floor to the intended speaker (Bohus & Horvitz, 2010).

Seeing an agent perform sensible eye gaze during conversation improves people's perceptions of that agent. When a robot is a listener in a multi-party conversation, seeing the robot track the conversation with its gaze elicits higher evaluations of that robot's comprehension and naturalness than seeing the robot perform random gaze turns between speakers (Kousidis & Schlangen, 2015). A robot that displays gaze focused on its human conversational partner but occasionally responds to motion in the background is evaluated as more natural, human-like, and attentive than a robot that exclusively focuses on the partner or that distributes its gaze randomly (Sorostinean, Ferland, Dang, & Tapus, 2014). Virtual avatars that use turn taking gaze during conversations are evaluated as more natural and more pleasant, and their conversation is rated as more engaging, than avatars that use random gaze or no gaze in their communication (Garau, Slater, Bee, & Sasse, 2001). In an immersive virtual reality setting, researchers confirmed that people have more positive subjective evaluations of an agent when it performs conversationally-driven gaze than when it performs random gaze, but that the effect depends on the agent's appearance. More realistic avatars benefit from appropriate conversational gaze, but low-realism avatars, such as stick figures, are adversely affected by human-like gaze behavior (Garau et al., 2003).

## 5.2 Narration

Unlike conversation, narration primarily involves a single speaker. There may be a single listener or an audience with multiple listeners. Contexts that involve narration include lecturing (as with robot tutors providing information about a topic), storytelling (as with entertainment robots), and presenting (as with robot tour guides that describe museum exhibits). Challenges in narration involve ensuring information recall and directing attention to external information sources, such as exhibits in a museum.

The type of robot gaze performed during narration can influence how much information is remembered by listeners. Longer participant-directed gaze from a storytelling robot leads to better recall of story content (Mutlu, Forlizzi, & Hodgins, 2006). In contrast, virtual agent tutors that display more gaze toward the subject matter than toward their listener generate better retention of information (Andrist, Pejisa, Mutlu, & Gleicher, 2012a). In general, however, socially communicative gaze is better for ensuring information recall than no gaze or gaze that is incongruous with communicative goals (Huang & Mutlu, 2013). Gaze behaviors can also be combined with other socially supportive behaviors, such as natural gestures and empathetic facial expressions, to improve student performance in language learning from a robotic tutor (Saerbeck, Schut, Bartneck, & Janse, 2010).

For robot tutors, acquiring, monitoring, and maintaining user engagement are particularly important, because reduced engagement means reduced learning. Animated pedagogical agents can use gaze to maintain student engagement during tutoring (Johnson, Rickel, & Lester, 2000). When



diminishing attention is detected, robots (Szafr & Mutlu, 2012) and virtual tutors (D’Mello, Olney, Williams, & Hays, 2012) can use verbal and nonverbal cues, including gaze, to restore the listener’s attention. Reorienting student attention in response to diminished engagement increases information recall (Szafr & Mutlu, 2012), specifically on questions that require deep reasoning (D’Mello, Olney, Williams, & Hays, 2012).

Listener-directed robot gaze during tutoring and storytelling is correlated with positive perceptions of a robot. Subjective ratings of likability and other positive attributes are higher for robots that display more affiliative gaze (that is, gaze directed at the listener) than referential gaze (Andrist, Pejosa, Mutlu, & Gleicher, 2012a). Robots exhibiting gaze that correlates to the content of their communication are seen as more natural and competent (Huang & Mutlu, 2013), and longer gazes toward a listener yield greater feelings of likeability (Karreman, Sepúlveda Bradford, Dijk, Lohse, & Evers, 2013). Mutual gaze, when presented with other social behaviors like head nods and posture mimicking, greatly improves people’s perceptions of rapport with a virtual agent (Wang & Gratch, 2010). Joint attention from a robot toward the topic of discussion is seen as more human-like than only mutual gaze (Karreman, Sepúlveda Bradford, Dijk, Lohse, & Evers, 2013).

However, there are cases in which listener-directed gaze negatively impacts people’s perceptions of a robot or virtual agent. High levels of mutual gaze without other social behaviors can decrease rapport with a human user to the same levels as a virtual agent specifically designed to show boredom (Wang & Gratch, 2010). Additionally, the benefit of listener-directed gaze may be influenced by gender; when listening to a storytelling robot, men evaluate the robot more positively when it looks at them more frequently than at their partner, while women show the opposite effect (Mutlu, Forlizzi, & Hodgins, 2006).

Some presentations, such as guided tours, involve narration about material that is situated externally to the agent. Tour guide robots might present a new technology to a user (Sidner, Kidd, Lee, & Lesh, 2004), provide route directions (Ono, Imai, & Ishiguro, 2001; Satake et al., 2010), or give location tours of indoor spaces (Knight & Simmons, 2013). A primary challenge for this kind of narration is to direct attention toward objects of interest in the environment, which can be accomplished using deictic gaze, either alone or in combination with other nonverbal behaviors like gestures (Ono, Imai, & Ishiguro, 2001).

A tour guide’s deictic gaze has a positive effect on listener engagement and attention. When a robot uses deictic gaze in addition to spoken object references, people are more engaged, spending more time interacting with the robot and displaying more coordinated gaze behaviors than when the robot simply speaks without supportive gaze (Sidner, Kidd, Lee, & Lesh, 2004). When a robot displays deictic gaze that reflects the subject of its speech, people display more nodding and mutual gaze, signaling increased engagement, than they do when the robot’s deictic gaze occurs at random points in its speech (Kuno et al., 2007; Yamazaki, Yamazaki, Burdelski, Kuno, & Fukushima, 2010). When listening to a robot tour guide, listener gaze directed away from the robot is often congruent with the robot’s topic of discussion (Knight & Simmons, 2013), indicating that robots can successfully guide listener attention to desired locations.

Tour guides can affect a listener’s experience by whether they look at the listener or at the display. A robot that orients its body (including its eyes) toward an exhibit can more easily engage its listeners than a robot that orients its eyes toward the audience, but people lose interest in the robot and its narrative more often when the robot looks at the exhibit and not at its audience (Karreman, Ludden, Dijk, & Evers, 2015). Robots can influence people’s experience of a tour by how often they direct gaze to each listener. When a robot “favors” a person by gazing at them longer than others in the group, that person reports greater feelings of likability toward the robot (Karreman, Sepúlveda Bradford, Dijk, Lohse, & Evers, 2013).

### 5.3 Collaboration

Collaboration requires communication of goals, knowledge, and intentions. For example, a robot that helps a user construct furniture needs to express its current goals and intended action to fluidly collaborate with a human partner. Gaze can be used to reveal these mental states to a partner in unobtrusive ways. Collaboration often involves the physical environment, so in addition to gaze that reveals mental states, such interactions also require gaze that references objects and physical locations.

Revealing mental states through nonverbal communication (including eye gaze) makes cooperative task performance faster, with errors detected more quickly and handled more effectively than purely task-based nonverbal communication (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005). Indicating engagement and providing feedback through subtle gaze behaviors improves performance of a human-robot team (Jung et al., 2013). Users also report understanding the robot better during their collaboration when it makes its mental models explicit (Breazeal, Kidd, Thomaz, Hoffman, & Berlin, 2005). Expressive eye gaze is one behavior (among many drawn from animation principles) that can make intentions and desires more explicit, for instance, by looking at a door handle when wanting to open a door (Takayama, Dooley, & Ju, 2011). Even when users are unaware of the intended communication, robots can “leak” their intentions through eye gaze, influencing human behavior in measurable ways (Mutlu, Yamaoka, Kanda, Ishiguro, & Hagita, 2009).

One part of collaboration involves referencing objects in the environment. Joint attention from a companion robot effectively draws a user’s attention to where the robot is looking (Yonezawa, Yamazoe, Utsumi, & Abe, 2007). Eye gaze can also act as a reinforcement of pointing gestures (Saupé & Mutlu, 2014). A robot can use eye gaze to support its speech in a cooperative object selection task, in which a human user needs to select an object referenced by the robot as quickly as possible (Admoni, Datsikas, & Scassellati, 2014; Boucher et al., 2012). People can recognize and respond to predictive eye gaze that indicates spatial references, completing the task faster than if they had been relying on the robot’s speech alone.

Errors in robot gaze hinder speech understanding, because people expect the robot’s gaze to indicate what the robot intends to verbally reference (Admoni, Datsikas, & Scassellati, 2014; Huang & Mutlu, 2012; Staudte & Crocker, 2009, 2011). For tasks that involve a light cognitive load (for instance, selecting the object referred to by the robot as quickly as possible), people recover quickly from errors in robot eye gaze and show no difference between incongruent gaze and having no gaze cues at all (Admoni, Datsikas, & Scassellati, 2014; Huang & Mutlu, 2012). However, in more cognitively demanding tasks (such as deciding whether a statement about the referenced object’s visual features is true or false by comparing features of the referenced object to other visible objects), incongruencies in a robot’s eye gaze and speech lead to diminished performance even over having no gaze at all (Staudte & Crocker, 2009, 2011).

To improve collaboration, users can teach skills to robots by performing demonstrations of those skills (Argall, Chernova, Veloso, & Browning, 2009). Robots can use gaze to establish joint attention when learning from such demonstrations and solicit feedback when uncertainty is high (Lockerd & Breazeal, 2004). When a robot student responds to joint attention by following the human teacher’s gaze, it better conveys the robot’s internal states and knowledge, which leads to more efficient teaching: fewer errors, faster recovery from errors, and less repetition of learned information (Huang & Thomaz, 2011). People also rate the robot as more natural and competent at its task when it engages in joint attention (Huang & Thomaz, 2011). People are sensitive to the robot’s mental state when they are teaching it and will adjust their behavior (in terms of pauses, speed, and magnitude of motions) to account for the robot’s visual attention (Pitsch, Vollmer, & Mühlhlig, 2013). When there are multiple robots to be taught, people are sensitive to each robot’s gaze behavior; they look longer and are more engaged in teaching robots that actively seek mutual gaze than robots that passively follow

the human's attention when it shifts to the other robots (Xu, Zhang, & Yu, 2013).

#### 5.4 Manipulation

One of the primary benefits of robots as physically embodied systems is their capability to physically manipulate objects in their environments. Many robot manipulators are still isolated in factories or other carefully controlled settings, but manipulation robots are increasingly required to operate in environments inhabited by people (Kemp, Edsinger, & Torres-Jara, 2007). For example, robot caregivers or office assistants must be able to pick up, carry, and hand over objects to assist their users. Object handovers particularly benefit from social robot gaze, because this type of manipulation depends on coordination with a partner.

While people are generally capable of performing successful handovers without much thought, the process of handing an object to another individual employs a series of subtle but important nonverbal cues, including eye gaze (Fig. 4). A decision tree built on empirical data of human-human handovers reveals that joint attention (attending to the same location or object) is important for coordinating a handover between two people, but that mutual gaze (where both people make eye contact with each other) is not (Strabala, Lee, Dragan, Forlizzi, & Srinivasa, 2012; Strabala et al., 2013). Robot-to-human handovers are improved when a robot monitors its partner's eye gaze for attention and engagement, only releasing the object when the user's focus of attention has turned to that object (Grigore, Eder, Pipe, Melhuish, & Leonards, 2013). In multi-party scenarios, robots can also use eye gaze to nonverbally select a member of the crowd to whom to hand an object (Kirchner, Alempijevic, & Dissanayake, 2011).

Gaze improves the efficiency of handovers. During a handover, people begin reaching for an object earlier when a robot continuously looks at the projected position in space where the handover will occur than when it looks away from that location (Moon et al., 2014). People reach for the object even earlier when the robot continually gazes at their faces than when it looks at the handover location (Zheng, Moon, Croft, & Meng, 2015). Gazes that transition between the user's face and handover location do not improve how quickly reaching begins, though people report that these gazes communicate the handover timing more effectively than continuous gazes (Zheng, Moon, Croft, & Meng, 2015).

Occasionally, a robot will need to direct a person's object manipulation, for example, when requesting that a person move an object within the robot's reach. Social behaviors including gaze cues can help inform people about where and how the robot would like such assistance; for example, the robot can look at the location to which it wants the object moved (Pandey, Ali, & Alami, 2013). This kind of social and referential gaze may be ignored during handovers, unless the receiver has good reason to interpret the robot's eye gaze as intentional (Admoni, Dragan, Srinivasa, & Scassellati, 2014). By introducing a social action—a delay in releasing the object—into the handover, robots can increase the amount of time spent looking at their gaze and user's compliance with gaze-based spatial references (Admoni, Dragan, Srinivasa, & Scassellati, 2014).

#### 5.5 Expression

Robots may benefit from the ability to express personality or emotion. For example, robots engaged in long-term interactions should have engaging personalities that keep the interactions from becoming stale; entertainment and companion robots may wish to express emotions that engage their users. Eye gaze is one way to express personality and emotion, though the challenge lies in generating the right kind of gaze to influence this subjective judgment.

Gaze behavior—in terms of where and for how long the robot gazes—can be used to express recognizable personalities and emotions. High levels of mutual gaze express feelings of trust (Normoyle et al., 2013) and extroversion (Andrist, Mutlu, & Tapus, 2015). Conversely, gaze aversions express

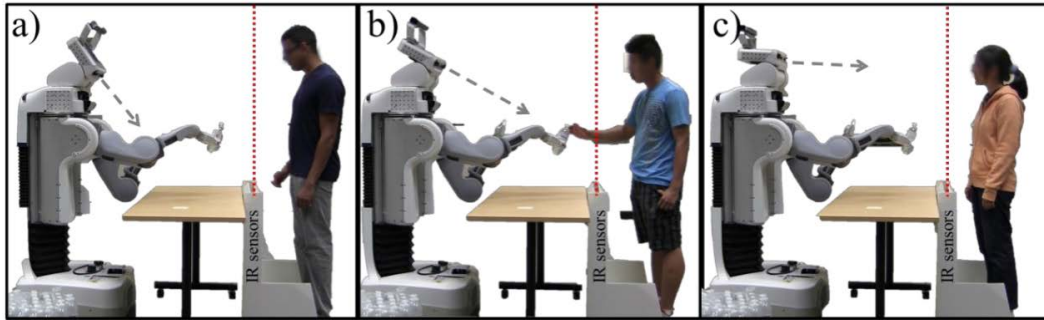


Figure 4. Gaze is an important part of successful handovers. Where a robot looks during a handover can influence how quickly people begin reaching for the object (Moon et al., 2014).

feelings of distrust (Normoyle et al., 2013) and introversion (Andrist, Mutlu, & Tapus, 2015). In animated agents, eye movement can be used to express recognizable emotions, such as joy, sadness, anger, fear, disgust, and surprise (Li & Mao, 2012b). Systematically manipulating features of gaze (such as amount of gaze, duration of gaze, and the points of fixation during gaze aversion) yields consistent impressions of dominance and friendliness in a robot (Fukayama, Ohno, Mukawa, Sawaki, & Hagita, 2002).

Robot expressiveness is important because it influences how people respond to the robot. Matching the robot's behavior with a user's personality (as evaluated by a personality survey) leads to greater motivation to engage in a repetitive task and improves subjective perceptions of a robot during collaboration (Andrist, Mutlu, & Tapus, 2015). When people perceive a robot favorably, they show no difference in proxemic behaviors when the robot increases its amount of mutual gaze. However, for people who dislike the robot, an increase in the robot's mutual gaze causes them to physically distance themselves from the robot (Mumm & Mutlu, 2011). Interestingly, this effect of interpersonal dynamics does not extend to psychological distancing during conversation, as measured by people's willingness to answer a series of revealing questions (Mumm & Mutlu, 2011).

## 5.6 Design-focused research summary

Design-focused studies have established that robot gaze can be used to improve human-robot interactions in a variety of domains. Robots can use a combination of mutual gaze and gaze aversions to regulate pace and participation in conversations, though the right amount and direction of gaze depends on the content of the conversation. Robot gaze can be used for deictic references; combining verbal and gaze-based deictic references leads to faster task performance than conveying information through speech alone. Robots can use their eye gaze to convey mental states, and doing so improves cooperation and learning. Gaze can also express personality and emotion that can improve rapport with users.

One unifying theme across these studies is that socially and contextually contingent gaze is more effective than gaze behaviors that are uncorrelated with the interaction. People respond more positively to robots, better remember discussion topics, and complete tasks more quickly when the robot's gaze is tied to what is being said or done. For example, people evaluate robots more highly when their gaze tracks the speakers in a conversation. Gaze directed at human partners also seems to have a positive effect on information recall and on how efficiently those partners complete cooperative tasks like handovers.

## 6. Technology-Focused Research

There are many approaches to achieving communicative social gaze from robots and virtual agents. One approach, which is grounded in the science of human cognition, models the underlying neurological or psychological processes of eye gaze (Section 6.1). The premise of this biologically inspired approach is that mimicking biology is an effective way to generate gaze that appears natural. Another approach is driven by behavioral data, basing gaze behaviors off of empirical measurements of gaze features—such as the timings, frequencies, and locations of gaze aversions—which are recorded during observations of human interactions (Section 6.2). This data-driven approach uses somewhat higher-level observations than the biologically inspired approach but still aims to replicate observed human behavior. A third approach is to construct heuristic systems that are not grounded in biological or empirical observations but (as with rules drawn from animation principles) still appear to generate expressive gaze (Section 6.3). Some systems cut across multiple approaches, such as a gaze model that uses observations of human behavior to derive heuristics, but in this section, we report studies based on their primary approach.

### 6.1 Biologically inspired systems

Biologically inspired gaze models attempt to replicate the underlying cognitive or neurobiological mechanisms that control gaze behavior in people. The systems in this category adhere to what we understand of the brain's function, though they operate at varying levels of detail. Some systems replicate the neuron-level receptive fields in the visual cortex to perform visual attention (Itti & Koch, 2001). Others employ a developmental approach to learning joint attention and other gaze behaviors; developmental robotics mimics human cognitive growth by attempting to replicate the process of human ontogenetic development (Lungarella & Metta, 2003). There is some evidence that biologically inspired models have higher accuracies than other gaze models in predicting human-like gaze fixations (Borji & Itti, 2013).

Many biologically inspired gaze models focus on directing attention to areas of interest in a visual scene by replicating the neurological response to those visual stimuli. These models generally have a similar structure: They compute the saliencies of several features in parallel, then combine these saliencies into a single saliency map (Frintrop, Rome, & Christensen, 2010). Both low-level scene features (e.g., color, intensity, and orientation) and high-level contextual features (e.g., object or scene recognition) are used to create saliency maps (Itti & Koch, 2001). Neurobiological models of gaze behavior that use bottom-up saliency maps can successfully track salient targets and perform visual search in demanding scenes (Itti & Koch, 2000), leading to realistic visual attention behavior in virtual avatars (Itti, Dhavale, & Pighin, 2004, 2006). Adding high-level contextual or motivational information to the low-level saliency cues enables robots to naturally direct their visual attention based on the current task and environment (Breazeal & Scassellati, 1999a). Behavior can be influenced by both visual and auditory saliency maps, as in an implementation of attention for the iCub robot (Ruesch et al., 2008). Other models of overt visual attention produce head and eye turns using dynamical neural networks that respond to visual saliency (Vijayakumar, Conradt, Shibata, & Schaal, 2001).

Cognitive models attempt to replicate high-level human cognition, so they operate at a level of abstraction above neuronal responses. Visual attention can be incorporated into cognitive models, such as ACT-R, to address how people's cognitive systems respond to environmental input (Anderson, Matessa, & Lebiere, 1997). For example, a computational cognitive architecture called ACT-R/E (ACT-R with Embodiment) performs conversational tracking by switching its visual attention to the speaker in a multi-party conversation (Trafton & Bugajska, 2008). By tightly integrating gaze behaviors with the underlying cognitive model controlling reasoning, dialogue management, and goals, the Rickel Gaze Model can generate real-time gaze shifts that reveal a virtual agent's internal

processes (Lee, Marsella, Traum, Gratch, & Lance, 2007).

Developmental models are also inspired by biology, though they attempt to replicate the higher-level cognitive process of learning rather than the underlying neuronal structure. For example, a computer vision model on a developmental robot uses saliency maps of the environment along with a probabilistic algorithm that estimates a teacher's gaze vectors to perform shared attention and gaze imitation; shared attention and imitation are foundational skills that bootstrap cognitive learning (Hoffman, Grimes, Shon, & Rao, 2006). A robot can develop the ability to perform joint attention through demonstrations of attention to salient objects, much in the same way that infants acquire this capability by interacting with their adult caregivers (Doniec, Sun, & Scassellati, 2006; Kozima & Yano, 2001; Nagai, Hosoda, Morita, & Asada, 2003; Triesch, Teuscher, Deák, & Carlson, 2006). These basic joint attention behaviors serve as the basis for learning more complex social communication skills in a humanoid robot (Scassellati, 1999).

## 6.2 Data-driven systems

The data-driven approach to generating robot eye gaze takes advantage of people's natural expressiveness by using quantified observations of human behavior to develop and train gaze systems. Though these systems use empirical behavioral data, they generally do not consider the underlying biological or cognitive mechanisms. The process of building data-driven systems generally follows three steps: First, observations of people using eye gaze in a desired scenario (such as in conversation) are collected. Second, a model of gaze behavior is developed from the gaze data in these observations, which are acquired either by manual coding or through automated feature extraction. Third, the behavior model is evaluated in a human-robot or human-agent interaction.

Data-driven researchers have recorded and analyzed human gaze behavior in a wide variety of scenarios. Conversational gaze has been recorded for pairs of previously unacquainted people discussing movie preferences (Andrist, Mutlu, & Gleicher, 2013), free dialogue between two people with various existing relationships (including hierarchical work relationships and romantic relationships) (Ishi, Liu, Ishiguro, & Hagita, 2010), and in four-person conversations about controversial topics such as "should euthanasia be legitimized" (Otsuka, Takemae, & Yamato, 2005). Observations of gaze in tutoring scenarios have been collected for student-teacher pairs covering topics as varied as paper making (Huang & Mutlu, 2014), board games (Admoni & Scassellati, 2014), and preparing canapés (Rich & Ponsler, 2010). Gaze data during object manipulation have been collected for individuals constructing Lego objects (Sakita, Ogawara, Murakami, Kawamura, & Ikeuchi, 2004).

Once the observational data are collected, they are annotated and processed to build a model of eye gaze within the specified interaction. Some models are built to generate robot behavior by extracting the features of gaze behaviors that achieve certain communicative functions. For example, researchers have extracted statistical information on timings and directions of gazes in dyadic conversations that achieve certain conversational functions, such as mediating turn-taking and regulating intimacy (Andrist, Mutlu, & Gleicher, 2013; Pelachaud & Bilvi, 2003). Others have identified the direction and timing of gaze during head tilts and nods in conversation (Liu, Ishi, Ishiguro, & Hagita, 2012) or during a physical construction task in which assistance may be required (Sakita, Ogawara, Murakami, Kawamura, & Ikeuchi, 2004). Models of gaze for narration can also incorporate other communicative behaviors, like gesture and speech (Huang & Mutlu, 2014).

After these models are developed, researchers must test their performance. For models that identify gaze behaviors to achieve certain communicative effects, researchers can incorporate these models into robot behavior generation systems and evaluate them in human-robot interactions. Data-driven gaze aversions in conversation lead to more disclosure from humans, better turn-taking regulation, and more positive subjective perceptions for virtual agents (Andrist, Mutlu, & Gleicher,

2013) and robots (Andrist, Tan, Gleicher, & Mutlu, 2014). Gaze during head tilts and nods that is generated according to a data-driven model increases the naturalness of a conversational robot (Liu, Ishi, Ishiguro, & Hagita, 2012). Robots that use a data-driven model to generate gazes and gestures during narration perform as well as robots that use hand-scripted behaviors (Huang & Mutlu, 2014).

In addition to models that generate gaze behavior, models built on gaze data can also provide information about interactions themselves, which can later be incorporated into other generative models. For example, a probabilistic model for multi-party conversation trained on gaze patterns during human-human interactions can identify the roles of participants in a conversation (Otsuka, Take-mae, & Yamato, 2005). A computational model trained on physical task tutoring data (constructing canapés or teaching about tangrams) can recognize gaze-based connection events that facilitate engagement between student and teacher (Holroyd, Rich, Sidner, & Ponsler, 2011; Rich & Ponsler, 2010). In a different tutoring interaction, a model of eye gaze and gesture trained with the k-nearest neighbor algorithm can predict what type of information the speaker is trying to communicate based on observations of nonverbal behaviors including gaze (Admoni & Scassellati, 2014). Because these models do not generate gaze behaviors directly, they are not evaluated through human-robot interactions. Instead, these models are evaluated by comparing their accuracy to ground-truth data that is annotated by humans.

### 6.3 Heuristic systems

A third approach to developing gaze technology employs heuristics that lead to appropriate looking behavior, regardless of actual biological function or human behavior. These heuristics allow researchers to directly design gaze behaviors, using understanding of psychology or knowledge of multimodal behavior, without being tied to underlying biological realities or requiring a large corpus of observational data.

One heuristic for generating gaze behaviors is to link a robot’s gaze to its speech. By representing each “communicative act” as comprised of a meaning (the information to transmit) and a signal (the nonverbal expression of that meaning), gaze can be closely integrated into the content of a robot’s speech (Poggi, Pelachaud, & De Rosis, 2000). A tool that automatically extracts syntactic and semantic information from a typed sentence can use that information to generate appropriate gaze behavior for a conversational virtual agent (Cassell, Vilhjálmsón, & Bickmore, 2004).

Gaze generation based on speech may not even require semantic understanding of that speech. Some social contexts can be extracted exclusively from the timing and structure of speech; using this information, a robot can automatically generate natural gaze behaviors that support the intended context without needing to understand what is being said (Murphy, Gonzales, & Srinivasan, 2011; Srinivasan & Murphy, 2011). Even loose coordination between a robot’s head motions and the sentence structure of its intended speech leads to reasonable, socially acceptable gazes in tele-operated or Wizard of Oz settings (Srinivasan, 2014).

Another heuristic for generating gaze behavior is to respond directly to a user’s gaze. For example, one human-aware manipulation planner for robot-to-human handovers takes into account where people are looking to inform where the handover should take place, and then communicates the robot’s intention to perform the handover by having the robot look at the object to be given (Sisbot & Alami, 2012). A robot behavior system for collaboration is responsive to fine-grained, real-time human eye movements collected with a head-mounted eye tracker (Xu, Zhang, & Yu, 2013). A virtual agent for conversation monitors a user’s gaze to assess their level of interest and responds with pre-specified gaze behaviors to elicit and maintain the user’s engagement (Peters, Pelachaud, Bevacqua, Mancini, & Poggi, 2005). Gandalf, an embodied conversational agent that teaches people about the solar system, detects users’ gaze acts and generates its own gaze in response to support its lesson (Thórisson, 1997,?). Systems that detect and respond to eye gaze can be used to shape

behaviors, for instance to promote social skills like joint attention to children with ASD (Bekele et al., 2013; De Silva, Tadano, Saito, Lambacher, & Higashi, 2009).

Such responsive systems can also account for multimodal inputs that include auditory or gesture information in addition to gaze. Combining auditory cues like sound source localization with visual cues like face detection, one robot performs mutual gaze and joint attention with viewers while presenting a museum exhibit (Bennewitz, Faber, Joho, Schreiber, & Behnke, 2005). Another robot can take human gaze direction, deictic gestures, and mood into account to attend to and interact with multiple people simultaneously (Spexard, Hanheide, & Sagerer, 2007).

A major source of heuristics for gaze behaviors is the psychology literature. Models built with heuristics drawn from psychology do not attempt to precisely replicate known cognitive functions. Moreover, unlike data-driven models, which observe human behavior in the precise task to be performed by a robot, heuristics drawn from psychology are not specific to a single scenario. For example, using approximate timings of face-directed and averted gaze from the psychology literature, as well as from informal observations, gaze behavior systems can support real-time conversation with virtual agents (Colburn, Cohen, & Drucker, 2000), as well as expressions of emotion and responses to environmental distractions (Gu & Badler, 2006). The Automated Visual Attending system uses rules drawn from psychology to generate attention behaviors in a virtual agent, in which goal-oriented intentional behaviors compete with involuntary attentional responses to stimuli (Khullar & Badler, 2001). A parametric computational model for animating gaze shifts of virtual agents that uses features informed by neurophysiology is successful at performing gaze shifts to peripheral targets (Andrist, Pejsa, Mutlu, & Gleicher, 2012b). Using a psychologically-based emotional model called the Geneva Emotion Wheel, a virtual agent expresses primary and secondary emotions by drawing pre-defined movement parameters for each emotional expression (Li & Mao, 2012a, 2012b). In multi-party interactions, visual attention on a very realistic humanoid robot is driven by a context-dependent social gaze generation system that accounts for multimodal features, such as proxemics, field of view, and verbal and nonverbal cues from the environment (Zaraki, Mazzei, Giuliani, & De Rossi, 2014).

#### 6.4 Tech-focused research summary

Approaches to generating social gaze in robots extend from biologically-based models to empirical models to heuristic systems that capture high-level rules of gaze behaviors. All of these approaches have been successful at generating gaze that improves interactions, but each of these approaches has drawbacks as well. Which approach to use for a technology-focused contribution depends on how important it is for gaze to be grounded in realistic behaviors versus how important it is to have design control over the behaviors.

Biological models typically focus on the parts of the neural system psychologists understand well, such as the visual attention system. Cognitive architectures attempt to produce more complex gaze behavior, but the behavior emerges from the structure of the system and cannot be precisely designed. Empirical systems require a data-collection phase that can be time consuming, but they result in gaze behaviors that perform as well or better than hand-tuned systems. The benefit of having gaze behaviors grounded in empirical data has to be weighed against the cost of collecting and annotating these data. Heuristic systems allow designers to more fully specify how gaze behaviors appear, but these designed behaviors may not correspond to the way gaze is actually used in human interactions.

## 7. Open Research Questions

This section presents open research questions about social eye gaze for human-robot interaction and discusses how each might be investigated. The first question asks how important it is for a robot



to have physical gaze capabilities, such as independently movable eyes and dilating pupils (Section 7.1). The second question addresses the impact of social agency observed throughout studies of gaze in HRI, particularly in differences between low-level and high-level responses to robot gaze (Section 7.2). The third question explores how physical embodiment affects the interpretation of a robot's eye gaze (Section 7.3). The final question addresses how gaze can be situated as an element of multimodal communication that includes other verbal and nonverbal behaviors (Section 7.4).

### 7.1 What is the role of physical capability in eye gaze for HRI?

As discussed in Section 2.1, research on gaze in HRI is conducted on robots with a range of physical capabilities. These capabilities, which replicate the subtle effects of human gaze—such as pupil dilation, saccades, and expressive secondary features like eyebrows—can provide additional social cues during interaction, but they are difficult to implement on physical systems.

The role of these capabilities has not yet been fully characterized. For example, many of the robots currently used for HRI research (such as the Nao) have fixed eyes and must move their entire head to indicate gaze shifts (Fig. 1). Robots like Keepon take this restriction a step further, requiring entire body shifts to indicate gaze direction. But head movements might be insufficient to communicate more subtle or rapid gazes. There is evidence that simply estimating a person's head pose from an RGB-D image is not enough to reliably ascertain their gaze direction (Kennedy, Baxter, & Belpaeme, 2015b), likely because people orient to lateral visual targets with saccades as well as head turns (Freedman & Sparks, 2000). It is not clear what information is lost when robots do not have the same capability of independent eye motion. However, implementing independently articulated eyes is costly (Section 2.1), so the value of robot eye gaze capabilities has to be clear before these capabilities are widely adopted.

Mapping gaze behavior from virtual agents, which have nearly unlimited capabilities, to physical robots, which are constrained by hardware, is not trivial (Ruhland et al., 2015). Understanding the effect of each capability will allow researchers to avoid over-generalizing their findings from virtual agents to embodied robot interactions. It will also enable robot designers to selectively implement hardware capabilities for specific effects, minimizing robot costs and complexity.

Identifying the importance of physical gaze capability for HRI is primarily a design-focused research question, because it addresses how modifying the physical appearance and behavior of a robot can affect interactions with people. Human-focused research may also contribute to this question, because characterizing the role of robot gaze might involve insights into characteristics and limits of human perception. Technology-focused research that aims to implement various gaze capabilities (for instance, saccades or eye blinks) would enable researchers to further investigate biologically realistic eye gaze in HRI.

### 7.2 What is the difference between micro-scale and macro-scale responses to robot gaze?

People's behavior can be measured at different levels. Micro-scale behaviors, such as eye saccades and reflexes, occur rapidly (within hundreds of milliseconds) and often over small distances (such as tiny shifts of gaze). Macro-scale behaviors are measured over larger times and distances, for instance, measurements of large gaze shifts for object references or observations of eye gaze over an entire interaction.

As described in Section 4, people's responses to robot gaze differ from their responses to human gaze in some ways when measured on a micro time scale. For instance, robot directional gaze does not cue the same reflexive response as human directional gaze in the first 500 milliseconds of exposure (Admoni, Bank, Tan, & Toneva, 2011). In the moments just before naming an object, people spend more time ensuring joint attention by looking at their partner's face than at the object if their partner is a robot, but more time looking at the object than at their partner's face if their partner

is another human (Yu, Schermerhorn, & Scheutz, 2012).

Conversely, there is much evidence that robot gaze has macro-scale behavioral effects that follow expected patterns from human-human interactions. For instance, robot gaze modeled after human behavior can successfully convey object references (Admoni, Datsikas, & Scassellati, 2014) and manage conversational turn-taking (Andrist, Tan, Gleicher, & Mutlu, 2014).

Researchers do not yet understand why these differences between micro- and macro-scale responses emerge. The differences may be artifacts of the experiments, they may arise from people's expectations of the robot, or they may have some other cause. Investigating the source of these differences is a human-focused research contribution that would enable researchers to develop robot gaze that has a specific, targeted effect on human behavior.

One explanation is that environmental cues affect how people respond to robot eye gaze. Micro-scale experiments are often well controlled because of the precision required to measure small changes of behavior. As described in section Section 2.3, these kinds of studies have the disadvantage of reduced ecological validity. It may be that these artificial settings affect people's natural responses to robots by reducing how much importance is attributed to their eye gaze.

Another possibility is that there exists a difference in automatic versus conscious processing of robots. People may not automatically attribute importance to robot eye gaze (as measured on a micro level, where responses tend to be reflexive), but context and their own expectations lead them to treat robot gaze as a meaningful stimulus when it is consciously processed (which can be seen in macro-scale measurements). Most studies in this review operate on the macro level and find that robot eye gaze has an effect on human behavior. In contrast, the studies described in Section 4.2 focus on either micro-scale measurements or on investigations of infant behavior, both of which involve human responses performed rapidly and with little high-level cognitive control. On these levels, people seem to respond to robot gaze as though it has no *a priori* social significance.

This automatic processing may be manipulable. In reflexive cueing experiments (Section 4), for example, a single stimulus presented as a social image (a face) can cue reflexive attention shifts, while the same image presented as a non-social image (a car) fails to cue these reflexive responses (Ristic et al., 2005), just as robots failed to cue these reflexive responses in an HRI study (Admoni, Bank, Tan, & Toneva, 2011). Does top-down context influence how people process a robot's eye gaze? Specifically, could people's response to a robot's eye gaze be changed just by whether the experimenter indicated that the robot was a social agent?

Presently, the quantity of research on the macro-scale effects of robot gaze vastly outweighs the quantity of research on gaze's micro-scale effects. One reason is that measuring micro-scale effects requires carefully controlled environments and precise sensors, while macro-scale effects can be measured in more naturalistic settings with common tools like video cameras. Further investigations into the differences between micro-scale and macro-scale effects of robot gaze is warranted. Understanding the disparity between these effects would involve human-focused research that investigates human cognitive processing, which could lead to both insights into human cognition and contributions about how to design effective robots for social interaction.

### 7.3 Under what conditions is embodiment important for the success of a robot's gaze behavior?

Gaze in HRI has been explored using both virtual agents and physically embodied robots (Section 2.1). By virtue of being animated, virtual agents provide hypothetically unlimited realism in their gaze behaviors, which can be achieved, for example, by implementing complex biological models of gaze (Ruhland et al., 2015). However, as described in Section 2.2, interactions with physically embodied robots may result in different human performance than interactions with virtual robots or videos of robots.

Physically embodied robots have been shown to increase cognitive learning gains (Leyzberg,

Spaulding, Toneva, & Scassellati, 2012) and compliance with robot instructions (Bainbridge, Hart, Kim, & Scassellati, 2011), though this effect does not hold in all studies (Kennedy, Baxter, & Belpaeme, 2015a; Powers, Kiesler, Fussell, & Torrey, 2007). Robots with physically embodied, 3D eyes, like the iCub (Metta, Sandini, Vernon, Natale, & Nori, 2008) or FACE (Zaraki, Mazzei, Giuliani, & De Rossi, 2014), may also offer a better view of the robot’s gaze direction, allowing people to infer the robot’s gaze target from a range of viewing angles. However, performing social gaze through embodied eyes must be balanced with the need to capture stable images for computer vision algorithms. This often leads to a functional split, where robots acquire their vision from cameras that are separate from their expressive eyes.

Though researchers have investigated the effects of physical embodiment in cognitive tasks, there is little research on whether embodiment influences the effect of eye gaze, specifically, in human-robot interactions. We do not yet know under what conditions, if any, physical robot embodiment influences the processing of robot eye gaze. Because of this, we do not yet have evidence that research findings about eye gaze from virtual agents can be generalized to embodied robot interactions, and vice versa.

Questions regarding physical embodiment for gaze in HRI include: Is there a difference in the emotional expressivity of virtual and physical eyes? Do people feel attention from virtual eyes as they do from physically embodied eyes? Can embodied eyes communicate a robot’s internal states through subtle cues as effectively as virtual eyes? Answers to these questions would make a design-focused contribution to eye gaze in HRI and may reveal human-focused insights about cognitive processing of embodied gaze.

Robots with physical embodiment but animated eyes, such as Baxter, the IROMEC robot (Ferrari, Robins, & Dautenhahn, 2009), and Chester (Vázquez, Steinfeld, Hudson, & Forlizzi, 2014), present an interesting middle ground of physical embodiment with which to evaluate these open questions (Fig. 5). Differences in gaze effects between robots with virtual eyes and those with 3D eyes may help identify the specific effects of physical *eyes*, as opposed to physical *bodies*, when examining eye gaze in human-robot interactions.

One concern about virtual eyes is that flat, two-dimensional displays sometimes create a powerful illusion, commonly called the “Mona Lisa effect,” that the eyes on the display are following the viewer regardless of viewing angle (Al Moubayed, Edlund, & Beskow, 2012; Delaunay, Greeff, & Belpaeme, 2010). Because gaze direction is an important indicator of attention and spatial references, the Mona Lisa effect limits the ability for such systems to communicate. One approach to dispel this effect is to back-project a virtual face onto a contoured three-dimensional surface (Al Moubayed, Beskow, Skantze, & Granström, 2012; Delaunay, Greeff, & Belpaeme, 2009; Kuratate, Matsusaka, Pierce, & Cheng, 2011). Back-projected technology provides the flexibility of animated eyes with the appearance of a more embodied system, as described in Section 2.1.

#### 7.4 How can eye gaze be integrated with other social behaviors?

Eye gaze is generally not produced in a vacuum. Instead, gaze often supports and augments other social behaviors (e.g., speech and gestures). Mental states (e.g., goals or cognitive effort) can also influence gaze. This multimodal behavior weaves a rich tapestry of social communication, and it is important to characterize what part eye gaze plays in the whole. Many of the studies described in this review isolated gaze behaviors in a controlled way in order to understand the specific characteristics of gaze in a particular situation. Broader-scale investigations into the role of eye gaze as a component of multimodal behavior will produce design- and technology-focused contributions toward more robust behavior models capable of acting across contexts, particularly outside the controlled laboratory environment.

Since speech is such a dominant modality in human interactions, it is difficult to separate gaze



*Figure 5.* Robots with physical bodies but animated eyes provide an interesting edge case when exploring the effects of embodiment on eye gaze in HRI (Section 7.3).

from speech in natural human-robot interactions. Researchers have shown that eye gaze improves speech-based interactions like conversations (Section 5.1) and narration (Section 5.2) by disambiguating object references, guiding attention, maintaining engagement, mediating turn-taking, managing partners, and many other functions. Further investigating how eye gaze and speech are linked, and how gaze can improve natural language processing and production in robots, would create better holistic systems for human-robot interactions.

Similarly, gaze and gestures are complementary nonverbal behaviors that often support each other in human interactions (McNeill, 1992). Here, gestures can be considered to be the set of nonverbal behaviors that include arm motions, body postures, proxemics, facial expressions, and other actions produced with the body that communicate social information. Some researchers have investigated gaze and gestures in the same context (Huang & Mutlu, 2013; Mutlu, Forlizzi, & Hodgins, 2006; Saupé & Mutlu, 2014), but further research to characterize how gaze and gestures interact would enable richer nonverbal models of robot communication.

Gaze behavior does not just influence other communication modalities; it is also modulated itself by cognitive processes like attention, memory, and mental effort. For example, psychologists know that pupils will dilate with increased cognitive effort (Beatty & Lucero-Wagoner, 2000), and that blinking is correlated with the beginning and end of information processing (Siegle, Ichikawa, & Steinhauer, 2008). Expressive robots could take advantage of these fine-grained gaze behaviors to convey mental states—for example, when they are thinking, when they are waiting for a response, or when they are experiencing difficulty—in a natural and human-like way. One use case for such expressiveness is for systems that learn from demonstration (Argall, Chernova, Veloso, & Browning, 2009). Robots that provide gaze feedback about their mental states to human tutors lead to more efficient teaching (Section 5.3). Taking advantage of gaze as a subtle but communicative behavior would improve learning from demonstration and other human-robot interactions.

## 8. Acknowledgements

This work was supported by the National Science Foundation Grant #113907 (Expeditions in Computing: Socially Assistive Robots) and US Office of Naval Research Grant #N00014-12-1-0822 (Social Hierarchical Learning). The authors thank Nicole Salomons, Marynel Vázquez, and anonymous reviewers for their helpful comments, advice, and assistance.

## References

- Admoni, H., Bank, C., Tan, J., & Toneva, M. (2011). Robot gaze does not reflexively cue human attention. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society (CogSci)* (p. 1983-1988). Austin, TX: Cognitive Science Society.
- Admoni, H., Datsikas, C., & Scassellati, B. (2014). Speech and gaze conflicts in collaborative human-robot interactions. In *Proceedings of the Annual Conference of the Cognitive Science Society (CogSci)* (pp. 104–109). Quebec City, Canada.
- Admoni, H., Dragan, A., Srinivasa, S., & Scassellati, B. (2014). Deliberate delays during robot-to-human handovers improve compliance with gaze communication. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 49–56). Bielefeld, Germany.
- Admoni, H., Hayes, B., Feil-Seifer, D., Ullman, D., & Scassellati, B. (2013). Are you looking at me? Perception of robot attention is mediated by gaze type and group size. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 389–396). Tokyo, Japan.
- Admoni, H., & Scassellati, B. (2012). Robot gaze is different from human gaze: Evidence that robot gaze does not cue reflexive attention. In *Proceedings of the “Gaze in Human-Robot Interaction” Workshop at HRI 2012*. Boston, MA.
- Admoni, H., & Scassellati, B. (2014). Data-driven model of nonverbal behavior for socially assistive human-robot interactions. In *Proceedings of the ACM International Conference on Multimodal Interaction (ICMI)*. Istanbul, Turkey.
- Al Moubayed, S., & Skantze, G. (2012). Perception of gaze direction for situated interaction. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction (Gaze-In)* (p. 1-6). Santa Monica, CA.
- Aldebaran. (2015, July). [Website] Retrieved from: <https://www.aldebaran.com/en>.
- Al Moubayed, S., Beskow, J., Skantze, G., & Granström, B. (2012). Furhat: A back-projected human-like robot head for multiparty human-machine interaction. In *Cognitive Behavioural Systems* (Vol. 7403, pp. 114–130). Springer Berlin Heidelberg.
- Al Moubayed, S., Edlund, J., & Beskow, J. (2012, January). Taming Mona Lisa: Communicating gaze faithfully in 2d and 3d facial projections. *ACM Transactions on Interactive Intelligent Systems*, 1(2).
- Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4), 439–462.
- Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology*, 6(1016).
- Andrist, S., Mutlu, B., & Gleicher, M. (2013). Conversational gaze aversion for virtual agents. In E. R. Aylett, B. Krenn, C. Pelachaud, H. Shimodaira (Ed.), *Intelligent Virtual Agents* (Vol. LNCS 8108, pp. 249–262).
- Andrist, S., Mutlu, B., & Tapus, A. (2015, April). Look Like Me: Matching Robot Personality via Gaze to Increase Motivation. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)*. Seoul, Republic of Korea: ACM Press.
- Andrist, S., Pejsa, T., Mutlu, B., & Gleicher, M. (2012a). Designing effective gaze mechanisms for virtual agents. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (pp. 705–714). Austin, TX: ACM Press.
- Andrist, S., Pejsa, T., Mutlu, B., & Gleicher, M. (2012b). A head-eye coordination model for animating gaze shifts of virtual characters. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. Santa Monica, CA: ACM Press.
- Andrist, S., Tan, X. Z., Gleicher, M., & Mutlu, B. (2014). Conversational gaze aversion for humanlike robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Bielefeld, Germany: ACM.
- Argall, B. D., Chernova, S., Veloso, M., & Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and Autonomous Systems*, 57(5), 469–483.
- Argyle, M. (1972). Non-verbal communication in human social interaction. In R. A. Hinde (Ed.), *Non-verbal communication*. Oxford, England: Cambridge University Press.
- Argyle, M., & Cook, M. (1976). *Gaze and mutual gaze*. Oxford, England: Cambridge University Press.

- Argyle, M., & Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica*, 6(1), 32–49.
- Bailly, G., Raidt, S., & Elisei, F. (2010, June). Gaze, conversational agents and face-to-face communication. *Speech Communication*, 52(6), 598–612.
- Bainbridge, W. A., Hart, J. W., Kim, E. S., & Scassellati, B. (2011). The benefits of interactions with physically present robots over video-displayed agents. *International Journal of Social Robotics*, 3, 41–52.
- Bartneck, C., Bleeker, T., Bun, J., Fens, P., & Riet, L. (2010). The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *PALADYN Journal of Behavioral Robotics*, 1, 109–115.
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1, 71–81.
- Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (p. 142-162). Oxford, England: Cambridge University Press.
- Bekele, E., Lahiri, U., Swanson, A., Crittendon, J., Warren, Z., & Sarkar, N. (2013). A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 21(2), 289–299.
- Bennewitz, M., Faber, F., Joho, D., Schreiber, M., & Behnke, S. (2005, December). Towards a humanoid museum guide robot that interacts with multiple persons. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (pp. 418–423). Tukuba, Japan: IEEE.
- Bohus, D., & Horvitz, E. (2010). Facilitating multiparty dialog with gaze, gesture, and speech. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI)* (p. 1). Beijing, China.
- Borji, A., & Itti, L. (2013, January). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1), 185-207.
- Boucher, J.-D., Pattacini, U., Lelong, A., Bailly, G., Elisei, F., Fagel, S., et al. (2012, January). I reach faster when i see you look: Gaze effects in human-human and human-robot face-to-face cooperation. *Frontiers in Neurobotics*, 6(May), 1–11.
- Breazeal, C., Hoffman, G., & Lockerd, A. (2004). Teaching and working with robots as a collaboration. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (pp. 1030–1037). IEEE Computer Society.
- Breazeal, C., Kidd, C. D., Thomaz, A. L., Hoffman, G., & Berlin, M. (2005). Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 708–713.
- Breazeal, C., & Scassellati, B. (1999a). A context-dependent attention system for a social robot. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*. Stockholm, Sweden.
- Breazeal, C., & Scassellati, B. (1999b). How to build robots that make friends and influence people. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vol. 2, pp. 858–863). Kyongju, South Korea.
- Brooks, R., & Meltzoff, A. N. (2002). The importance of eyes: How infants interpret adult looking behavior. *Developmental Psychology*, 38(6), 958–966.
- Broz, F., Lehmann, H., Nehaniv, C. L., & Dautenhahn, K. (2012, March 5). Mutual gaze, personality, and familiarity: Dual eye-tracking during conversation. In *Proceedings of the “Gaze in Human-Robot Interaction” Workshop at the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Boston, MA.
- Bruce, A. (2002). The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (p. 38-42). Washington, DC: IEEE Press.
- Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., et al. (2002). Reading the mind from eye gaze. *Neuropsychologia*, 40(8), 1129–1138.
- Cappella, J. N., & Pelachaud, C. (2002). Rules for responsive robots: Using human interactions to build virtual interactions. *Stability and Change in Relationships*, 325–354.

- Cassell, J. (2000, April). Embodied conversational interface agents. *Communications of the ACM*, 34(4).
- Cassell, J., Torres, O., & Prevost, S. (1998). Turn taking vs. discourse structure: How best to model multimodal conversation. In *Machine Conversations* (pp. 143–154). Kluwer.
- Cassell, J., Vilhjálmsón, H. H., & Bickmore, T. (2004). BEAT: The behavior expression animation toolkit. *Life-Like Characters*, 477–486.
- Chawarska, K., & Shic, F. (2009). Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 39(12), 1663–1672.
- Chidambaram, V., Chiang, Y.-H., & Mutlu, B. (2012, March). Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 293–300). Boston, MA: ACM Press.
- Choi, J. J., Kim, Y., & Kwak, S. S. (2013, March). Have you ever lied?: The impacts of gaze avoidance on people’s perception of a robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 105–106). Tokyo, Japan.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Colburn, R. A., Cohen, M. F., & Drucker, S. M. (2000). *The role of eye gaze in avatar mediated conversational interfaces* (Vol. 81; Tech. Rep.). Microsoft Research.
- Cook, M. (1977). Gaze and mutual gaze in social encounters: How long—and when—we look others “in the eye” is one of the main signals in nonverbal communication. *American Scientist*, 65(3), 328–333.
- Dautenhahn, K., Nehaniv, C. L., Walters, M. L., Robins, B., Kose-Bagci, H., Mirza, N. A., et al. (2009, December). KASPAR—a minimally expressive humanoid robot for human-robot interaction research. *Applied Bionics and Biomechanics*, 6(3–4), 369–397.
- De Silva, P. R. S., Tadano, K., Saito, A., Lambacher, S. G., & Higashi, M. (2009). Therapeutic-assisted robot for children with autism. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (p. 3561–3567). St. Louis, MO.
- Delaunay, F. (2015). *A retro-projected robotic head for social human-robot interaction*. Unpublished doctoral dissertation, Plymouth University, Plymouth, England.
- Delaunay, F., Greeff, J. de, & Belpaeme, T. (2009). Towards retro-projected robot faces: An alternative to mechatronic and android faces. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (p. 306–311). Toyama, Japan.
- Delaunay, F., Greeff, J. de, & Belpaeme, T. (2010). A study of a retro-projected robotic face and its effectiveness for gaze reading by humans. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 39–44). Osaka, Japan: IEEE.
- D’Mello, S., Olney, A., Williams, C., & Hays, P. (2012, May). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, 70(5), 377–398.
- Doniec, M., Sun, G., & Scassellati, B. (2006, December). Active learning of joint attention. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (p. 34–39). Genova, Italy.
- Downing, P. E., Dodds, C. M., & Bray, D. (2004). Why does the gaze of others direct visual attention. *Visual Cognition*, 11(1), 71–79.
- Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baron-Cohen, S. (1999). Gaze perception triggers reflexive visuospatial orienting. *Visual Cognition*, 6(5), 509–540.
- Emery, N. (2000). The eyes have it: the neuroethology, function and evolution of social gaze. *Neuroscience & Biobehavioral Reviews*, 24, 581–604.
- Ferrari, E., Robins, B., & Dautenhahn, K. (2009). Therapeutic and educational objectives in robot assisted play for children with autism. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (p. 108–114). Toyama, Japan.
- Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003, March). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3–4), 143–166.
- Freedman, E. G., & Sparks, D. L. (2000). Coordination of the eyes and head: movement kinematics. *Experimental Brain Research*, 131(1), 22–32.
- Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional effects of counterpredictive gaze and arrow cues.

- Journal of Experimental Psychology: Human Perception and Performance*, 30(2), 319–329.
- Frintrop, S., Rome, E., & Christensen, H. I. (2010, January). Computational visual attention systems and their cognitive foundations: A survey. *ACM Transactions on Applied Perception*, 7(1), 6:1–6:39.
- Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, 133(4), 694–724.
- Fukayama, A., Ohno, T., Mukawa, N., Sawaki, M., & Hagita, N. (2002). Messages embedded in gaze of interface agents — impression management with agent's gaze. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (p. 41). New York, NY: ACM Press.
- Garau, M., Slater, M., Bee, S., & Sasse, M. (2001). The impact of eye gaze on communication using humanoid avatars. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (p. 309-316). Seattle, WA: ACM Press.
- Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (Vol. 5, p. 529). New York, NY: ACM Press.
- Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11), 419 - 429.
- Griffin, Z. M., & Bock, K. (2000, July). What the eyes say about speaking. *Psychological Science*, 11(4), 274–279.
- Grigore, E. C., Eder, K., Pipe, A. G., Melhuish, C., & Leonards, U. (2013, November). Joint action understanding improves robot-to-human object handover. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Tokyo, Japan: IEEE.
- Gu, E., & Badler, N. (2006). Visual attention and eye gaze during multiparty conversations with distractions. In J. Gratch, M. Young, R. Aylett, D. Ballin, & P. Olivier (Eds.), *Intelligent virtual agents* (Vol. 4133, pp. 193–204). Springer Berlin Heidelberg.
- Ham, J., Cuijpers, R. H., & Cabibihan, J.-J. (2015). Combining robotic persuasive strategies: The persuasive power of a storytelling robot that uses gazing and gestures. *International Journal of Social Robotics*, 7, 479–487.
- Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57, 596–615.
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4), 188–194.
- Hoffman, M. W., Grimes, D. B., Shon, A. P., & Rao, R. P. (2006). A probabilistic model of gaze imitation and shared attention. *Neural Networks*, 19, 299–310.
- Holroyd, A., Rich, C., Sidner, C. L., & Ponsler, B. (2011, July). Generating connection events for human-robot collaboration. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (p. 241-246). Atlanta, GA: IEEE.
- Hood, B. M., Willen, J. D., & Driver, J. (1998, March). Adult's eyes trigger shifts of visual attention in human infants. *Psychological Science*, 9(2).
- Huang, C.-M., & Mutlu, B. (2012). Robot behavior toolkit: Generating effective social behaviors for robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 25–32). Boston, MA.
- Huang, C.-M., & Mutlu, B. (2013, June). The repertoire of robot behavior: Designing social behaviors to support human-robot joint activity. *Journal of Human-Robot Interaction (JHRI)*, 2(2), 80–102.
- Huang, C.-M., & Mutlu, B. (2014). Learning-based modeling of multimodal behaviors for humanlike robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 57–64). Bielefeld, Germany: ACM.
- Huang, C.-M., & Mutlu, B. (2016). Anticipatory robot control for efficient human-robot collaboration. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 83-90). Christchurch, New Zealand.
- Huang, C.-M., & Thomaz, A. L. (2011). Effects of responding to, initiating and ensuring joint attention in human-robot interaction. In *Proceedings of the IEEE International Symposium on Robot and Human*



- Interactive Communication (RO-MAN)* (p. 65-71). Atlanta, GA.
- Hyönä, J., Tammola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology*, 48(3), 598–612.
- Imai, M., Kanda, T., Ono, T., Ishiguro, H., & Mase, K. (2002). Robot mediated round table: Analysis of the effect of robot's gaze. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (p. 411-416).
- Ishi, C. T., Liu, C., Ishiguro, H., & Hagita, N. (2010). Head motion during dialogue speech and nod timing control in humanoid robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 293–300). Osaka, Japan: IEEE.
- Itakura, S., Ishida, H., Kanda, T., Shimada, Y., Ishiguro, H., & Lee, K. (2008). How to build an intentional android: Infants' imitation of a robot's goal-directed actions. *Infancy*, 13(5), 519–532.
- Ito, A., Hayakawa, S., & Terada, T. (2004, September). Why robots need body for mind communication—An attempt of eye-contact between human and robot. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (p. 473-478). Kurashiki, Okayama Japan: IEEE.
- Itti, L., Dhavale, N., & Pighin, F. (2004). Realistic avatar eye and head animation using a neurobiological model of visual attention. In B. Bosacchi, D. B. Fogel, & J. C. Bezdek (Eds.), *Spie* (Vol. 5200, pp. 64–78).
- Itti, L., Dhavale, N., & Pighin, F. (2006). Photorealistic attention-based gaze animation. In *Proceedings of IEEE International Conference on Multimedia and Expo (ICME)* (pp. 521–524). Toronto, Canada.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 40(10–12), 1489–1506.
- Itti, L., & Koch, C. (2001, March). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2, 194–203.
- Johansson, R. S., Westling, G., Bäckström, A., & Flanagan, J. R. (2001, September). Eye-hand coordination in object manipulation. *The Journal of Neuroscience*, 21(17), 6917–6932.
- Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- Jung, M. F., Lee, J. J., DePalma, N., Adalgeirsson, S. O., Hinds, P. J., & Breazeal, C. (2013). Engaging robots: Easing complex human-robot teamwork using backchanneling. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)* (p. 1555-1566). San Antonio, TX.
- Kang, S., Gratch, J., & Sidner, C. (2012). Towards building a virtual counselor: Modeling nonverbal behavior during intimate self-disclosure. In Conitzer, Winikoff, Padgham, & van der Hoek (Eds.), *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)* (p. 4-8). Valencia, Spain.
- Kanngiesser, P., Itakura, S., Zhou, Y., Kanda, T., Ishiguro, H., & Hood, B. (2015). The role of social eye-gaze in children's and adults' ownership attributions to robotic agents in three cultures. *Interaction Studies*, 16(1), 1–28.
- Karreman, D. E., Ludden, G. D., Dijk, E. M. van, & Evers, V. (2015, April). How can a tour guide robot influence visitors' engagement, orientation and group formations? In M. Salem, A. Weiss, P. Baxter, & K. Dautenhahn (Eds.), *4th International Symposium on New Frontiers in Human-Robot Interaction*. Canterbury, UK.
- Karreman, D. E., Sepúlveda Bradford, G., Dijk, B. van, Lohse, M., & Evers, V. (2013). What happens when a robot favors someone? In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 157–158). Tokyo, Japan.
- Kemp, C. C., Edsinger, A., & Torres-Jara, E. (2007). Challenges for robot manipulation in human environments. *IEEE Robotics and Automation Magazine*, 14(1), 20.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta Psychologica*, 26(1), 22–63.
- Kennedy, J., Baxter, P., & Belpaeme, T. (2015a). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics (IJSR)*, 7(2), 293-308.

- Kennedy, J., Baxter, P., & Belpaeme, T. (2015b). Head pose estimation is an inadequate replacement for eye gaze in child-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI) Late Breaking Report* (p. 35-36). Portland, OR: ACM.
- Khullar, S. C., & Badler, N. I. (2001). Where to look? Automating Attending Behaviors of Virtual Human Characters. *Autonomous Agents and Multi-Agent Systems*, 4(1-2), 9-23.
- Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition*, 55, 269–271.
- Kirchner, N., Alempijevic, A., & Dissanayake, G. (2011). Nonverbal robot-group interaction using an imitated gaze cue. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 497). New York, NY: ACM Press.
- Kleinke, C. L. (1986, July). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1), 78–100.
- Knight, H., & Simmons, R. (2013, May). Estimating human interest and attention via gaze analysis. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*. Karlsruhe, Germany.
- Kobayashi, H., & Kohshima, S. (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. *Journal of Human Evolution*, 40(5), 419–435.
- Kousidis, S., & Schlangen, D. (2015). The power of a glance: Evaluating embodiment and turn-tracking strategies of an active robotic overhearer. In *Proceedings of the 2015 AAAI Spring Symposium: Turn-Taking and Coordination in Human-Machine Interaction* (p. 36-43). Palo Alto, CA: AAAI Press.
- Kozima, H., & Ito, A. (1998). Towards language acquisition by an attention-sharing robot. In D. Powers (Ed.), *Proceedings of the International Conference on New Methods in Language Processing and Computational Natural Language Learning (CoNLL)* (p. 245-246). Sydney, Australia.
- Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon: A playful robot for research, therapy, and entertainment. *International Journal of Social Robotics*, 1, 3–18.
- Kozima, H., & Yano, H. (2001). A robot that learns to communicate with human caregivers. In *Proceedings of the International Workshop on Epigenetic Robotics* (p. 47-52). Lund, Sweden.
- Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A., & Kuzuoka, H. (2007). Museum guide robot based on sociological interaction analysis. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (p. 1191-1194). San Jose, CA.
- Kuratate, T., Matsusaka, Y., Pierce, B., & Cheng, G. (2011). "mask-bot": A life-size robot head using talking head animation for human-robot communication. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (pp. 99–104). Bled, Slovenia.
- Land, M. F., & Hayhoe, M. (2001, January). In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26), 3559–65.
- Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition*, 6(5), 541–567.
- Lee, J., Marsella, S., Traum, D., Gratch, J., & Lance, B. (2007). The rickel gaze model: A window on the mind of a virtual human. In C. Pelachaud (Ed.), *Intelligent virtual agents* (Vol. LNAI 4722, pp. 296–303). SpringerBerlin-Heidelberg.
- Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006, October). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in humanrobot interaction. *International Journal of Human-Computer Studies*, 64(10), 962–973.
- Leyzberg, D., Spaulding, S., Toneva, M., & Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Conference of the Cognitive Science Society (CogSci)*. Sapporo, Japan.
- Li, & Mao, X. (2012a, May). EEMML: The emotional eye movement animation toolkit. *Multimedia Tools and Applications*, 60(1), 181-201.
- Li, & Mao, X. (2012b, October). Emotional eye movement generation based on Geneva Emotion Wheel for virtual agents. *Journal of Visual Languages & Computing*, 23(5), 299–310.
- Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent

- robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77, 23–37.
- Liu, C., Ishi, C. T., Ishiguro, H., & Hagita, N. (2012, March). Generation of nodding, head tilting and eye gazing for human-robot dialogue interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 285-292). Boston, MA: ACM Press.
- Lockerd, A., & Breazeal, C. (2004). Tutelage and socially guided robot learning. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vol. 4, p. 3475-3480).
- Lungarella, M., & Metta, G. (2003). Beyond gazing, pointing, and reaching: A survey of developmental robotics. In *Proceedings of the Joint IEEE International Conference on Developmental and Learning and on Epigenetic Robotics* (pp. 81–89).
- Matsusaka, Y., Fujie, S., & Kobayashi, T. (2001). Modeling of conversational strategy for the robot participating in the group conversation. In *Proceedings of the European Conference on Speech Communication and Technology (Eurospeech)* (Vol. 1, p. 2173-2176).
- Matsusaka, Y., Tojo, T., Kubota, S., Furukawa, K., Tamiya, D., Hayata, K., et al. (1999). Multi-person conversation via multi-modal interface: A robot who communicate with multi-user. In *Proceedings of the European Conference on Speech Communications and Technology (Eurospeech)* (p. 1723-1726).
- Mavridis, N. (2015). A review of verbal and non-verbal human-robot interactive communication. *Robotics and Autonomous Systems*, 63, 22–35.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: The University of Chicago Press.
- Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). “Social” robots are psychological agents for infants: A test of gaze following. *Neural Networks*, 23, 966–972.
- Metta, G., Sandini, G., Vernon, D., Natale, L., & Nori, F. (2008). The iCub humanoid robot: An open platform for research in embodied cognition. In *Proceedings of the Workshop on Performance Metrics for Intelligent Systems* (pp. 50–56). Gaithersburg, MD: ACM.
- Moon, A., Troniak, D. M., Gleeson, B., Pan, M. K. X. J., Zheng, M., Blumer, B. A., et al. (2014). Meet me where i’m gazing: How shared attention gaze affects human-robot handover timing. In *Proceedings of the 9th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 334–341). Bielefeld, Germany: ACM.
- Moore, C., & Dunham, P. (2014). *Joint attention: Its origins and role in development*. New York, NY: Psychology Press.
- Movellan, J. R., & Watson, J. S. (2002). The development of gaze following as a Bayesian systems identification problem. In *Proceedings of the IEEE International Conference on Development and Learning (ICDL)* (pp. 34–40). Cambridge, MA.
- Mumm, J., & Mutlu, B. (2011, March). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Lausanne, Switzerland.
- Murphy, R., Gonzales, J., & Srinivasan, V. (2011). Inferring social gaze from conversational structure and timing. *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 209–210.
- Mutlu, B., Forlizzi, J., & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids)* (pp. 518–523). Genova, Italy.
- Mutlu, B., Kanda, T., Forlizzi, J., Hodgins, J., & Ishiguro, H. (2012, January). Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems*, 1(2).
- Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H., & Hagita, N. (2009, March). Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 61–68). La Jolla, CA: ACM.
- Mutlu, B., Yamaoka, F., Kanda, T., Ishiguro, H., & Hagita, N. (2009, March). Nonverbal leakage in robots: Communication of intentions through seemingly unintentional behavior. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 69–76). La Jolla, CA: ACM.

- Nagai, Y., Hosoda, K., Morita, A., & Asada, M. (2003, December). A constructive model for the development of joint attention. *Connection Science*, 15(4), 211–229.
- Normoyle, A., Badler, J. B., Fan, T., Badler, N. I., Cassol, V. J., & Musse, S. R. (2013). Evaluating perceived trust from procedurally animated gaze. In *Proceedings of Motion on Games* (pp. 119:141–119:148). New York, NY: ACM.
- Oertel, C., Włodarczak, M., Edlund, J., Wagner, P., & Gustafson, J. (2012, September). Gaze patterns in turn-taking. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)* (Vol. 3, pp. 2243–2246). Portland, OR.
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013a). Infants understand the referential nature of human gaze but not robot gaze. *Journal of Experimental Child Psychology*, 116(1), 86–95.
- Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013b, August). The power of human gaze on infant learning. *Cognition*, 128(2), 127–33.
- Ono, T., Imai, M., & Ishiguro, H. (2001). A model of embodied communications with gestures between humans and robots. In *Proceedings of the Annual Conference of the Cognitive Science Society (CogSci)* (p. 732-737). Edinburgh, Scotland.
- Otsuka, K., Takemae, Y., & Yamato, J. (2005). A probabilistic inference of multiparty-conversation structure based on Markov-switching models of gaze patterns, head directions, and utterances. In *Proceedings of the International Conference on Multimodal Interfaces* (pp. 191–198). New York, NY, USA: ACM.
- Pandey, A. K., Ali, M., & Alami, R. (2013). Towards a task-aware proactive sociable robot based on multi-state perspective-taking. *International Journal of Social Robotics*, 5(2), 215–236.
- Pelachaud, C., & Bilvi, M. (2003). Modelling gaze behavior for conversational agents. In *Intelligent Virtual Agents (IVA)* (Vol. LNAI 2792, p. 93-100). Germany: Springer-Verlag.
- Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., & Poggi, I. (2005). A model of attention and interest using gaze behavior. In T. Panayiotopoulos, J. Gratch, R. Aylett, D. Ballin, P. Olivier, & T. Rist (Eds.), *Intelligent virtual agents* (Vol. 3661, p. 229-240). Springer Berlin Heidelberg.
- Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011, January). A non-verbal Turing test: Differentiating mind from machine in gaze-based social interaction. *PloS one*, 6(11), e27591.
- Pfeiffer-Lessmann, N., Pfeiffer, T., & Wachsmuth, I. (2012). An operational model of joint attention-timing of gaze patterns in interactions between humans and a virtual human. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the Annual Conference of the Cognitive Science Society (CogSci)* (p. 851-856). Austin, TX: Cognitive Science Society.
- Pierno, A., Becchio, C., Wall, M., Smith, A., Turella, L., & Castiello, U. (2006). When gaze turns into grasp. *Journal of Cognitive Neuroscience*, 18(12), 2130–2137.
- Pitsch, K., Vollmer, A.-L., & Mühlh, M. (2013). Robot feedback shapes the tutor’s presentation: How a robot’s online gaze strategies lead to micro-adaptation of the human’s conduct. *Interaction Studies*, 14(2), 268–296.
- Poggi, I., Pelachaud, C., & De Rosi, F. (2000). Eye communication in a conversational 3D synthetic agent. *AI Communications*, 13(3), 169–181.
- Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 145-152). Washington, DC.
- Rich, C., & Ponsler, B. (2010). Recognizing engagement in human-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 375-382). Osaka, Japan.
- Ristic, J., Mottson, L., Friesen, C. K., Iarocci, G., Burack, J. A., & Kingstone, A. (2005). Eyes are special but not for everyone: The case of autism. *Cognitive Brain Research*, 24, 715–718. (Short communication)
- Ruesch, J., Lopes, M., Bernardino, A., Hörnstein, J., Santos-Victor, J., & Pfeifer, R. (2008, May). Multimodal saliency-based bottom-up attention: A framework for the humanoid robot iCub. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)* (p. 962-967). Pasadena, CA.
- Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., et al. (2015). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception. *Computer Graphics Forum*, 1–28.
- Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: varying the

- degree of social supportive behavior of a robotic tutor. In *Proceedings of the International Conference on Human Factors in Computing Systems (CHI)* (p. 1613-1622). Atlanta, GA.
- Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H., & Hagita, N. (2007). Android as a telecommunication medium with a human-like presence. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 193-200). Washington, DC.
- Sakita, K., Ogawara, K., Murakami, S., Kawamura, K., & Ikeuchi, K. (2004). Flexible cooperation between human and robot by interpreting human intention from gaze information. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vol. 1).
- Satake, S., Kanda, T., Glas, D. F., Imai, M., Ishiguro, H., & Hagita, N. (2010). How to approach humans?-strategies for social robots to initiate interaction. *Journal of the Robotics Society of Japan*, 28(3), 327–337.
- Sauppé, A., & Mutlu, B. (2014). Robot deictics: How gesture and context shape referential communication. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction (HRI)* (p. 342-349). Bielefeld, Germany.
- Scassellati, B. (1996). Mechanisms of shared attention for a humanoid robot. In *Proceedings of the AAAI Fall Symposium on Embodied Cognition and Action* (Vol. 4, p. 21). Cambridge, MA.
- Scassellati, B. (1999). Imitation and mechanisms of joint attention: A developmental structure for building social skills on a humanoid robot. In *Computation For Metaphors, Analogy, and Agents* (Vol. 1562, p. 176-195). Springer Lecture Notes in Artificial Intelligence.
- Scassellati, B. (2007). How social robots will help us to diagnose, treat, and understand autism. In S. Thrun, R. Brooks, & H. Durrant-Whyte (Eds.), *Robotics research* (Vol. 28, p. 552-563). Springer Berlin-Heidelberg.
- Scassellati, B., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, 14, 275–294.
- Shah, J., & Breazeal, C. (2010). An empirical analysis of team coordination behaviors and action planning with application to human-robot teaming. *Human Factors*, 52(2), 234-245.
- Shic, F., Scassellati, B., Lin, D., & Chawarska, K. (2007, July). Measuring context: The gaze patterns of children with autism evaluated from the bottom-up. In *Proceedings of the IEEE International Conference on Development and Learning (ICDL)* (p. 70-75). London, UK.
- Sidner, C. L., Kidd, C. D., Lee, C., & Lesh, N. (2004). Where to look: A study of human-robot engagement. In *Proceedings of the International Conference on Intelligent User Interfaces (IUI '04)* (pp. 78–84). New York, NY: ACM.
- Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology*, 45(5), 679–687.
- Simmons, R., Makatchev, M., Kirby, R., Lee, M. K., Fanaswala, I., Browning, B., et al. (2011). Believable robot characters. *AI Magazine*, 32(4), 39–52.
- Sisbot, E. A., & Alami, R. (2012, October). A human-aware manipulation planner. *IEEE Transactions on Robotics*, 28(5), 1045–1057.
- Sorostinean, M., Ferland, F., Dang, T.-H.-H., & Tapus, A. (2014). Motion-oriented attention for a social gaze robot behavior. In *Proceedings of the International Conference on Social Robotics (ICSR)* (pp. 310–319). Sydney, Australia: Springer. (LNAI 8755)
- Spexard, T., Hanheide, M., & Sagerer, G. (2007, October). Human-oriented interaction with an anthropomorphic robot. *IEEE Transactions on Robotics*, 23(5), 852–862.
- Srinivasan, V. (2014). *High social acceptance of head gaze loosely synchronized with speech for social robots*. Unpublished doctoral dissertation, Texas A & M University.
- Srinivasan, V., & Murphy, R. (2011, March). A survey of social gaze. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (p. 253-254). Lausanne, Switzerland.
- Staudte, M., & Crocker, M. W. (2009, March). Visual attention in spoken human-robot interaction. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 77–84). La Jolla, CA: ACM Press.
- Staudte, M., & Crocker, M. W. (2011, August). Investigating joint attention mechanisms through spoken human-robot interaction. *Cognition*, 120, 268–291.

- Strabala, K., Lee, M. K., Dragan, A., Forlizzi, J., & Srinivasa, S. S. (2012, September). Learning the communication of intent prior to physical collaboration. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 968–973). IEEE.
- Strabala, K., Lee, M. K., Dragan, A., Forlizzi, J., Srinivasa, S. S., Cakmak, M., et al. (2013). Toward seamless human-robot handovers. *Journal of Human-Robot Interaction (JHRI)*, 2(1), 112–132.
- Szafir, D., & Mutlu, B. (2012, May). Pay attention! Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the ACM Annual Conference on Human Factors in Computing Systems (CHI)* (pp. 11–20). Austin, TX.
- Takayama, L., Dooley, D., & Ju, W. (2011, March). Expressing thought: Improving robot readability with animation principles. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 69–76). Lausanne, Switzerland: ACM Press.
- Tapus, A., Matarić, M., & Scassellati, B. (2007, March). Socially assistive robotics: The grand challenges in helping humans through social interaction. *IEEE Robotics and Automation Magazine*, 35–42.
- Tapus, A., Peca, A., Aly, A., Pop, C., Jisa, L., Pinte, S., et al. (2012). Children with autism social engagement in interaction with Nao, an imitative robot A series of single case experiments. *Interaction Studies*, 13(3), 315–347.
- Thórisson, K. (1997). Gandalf: An embodied humanoid capable of real-time multimodal dialogue with people. In *Proceedings of the First International Conference on Autonomous Agents (AGENTS)*. Marina Del Rey, CA: ACM.
- Thórisson, K. R. (1994, August). Face-to-face communication with computer agents. In *AAAI Spring Symposium on Believable Agents* (pp. 86–90). Stanford University, CA. Retrieved from [http://alumni.media.mit.edu/~kris/ftp/ft\\_f\\_comm.pdf](http://alumni.media.mit.edu/~kris/ftp/ft_f_comm.pdf)
- Thórisson, K. R. (1997). Layered modular action control for communicative humanoids. In *Computer Animation* (pp. 134–143).
- Trafton, J., & Bugajska, M. (2008). Integrating Vision and Audition within a Cognitive Architecture to Track Conversations. In *Proceedings of the 3rd International Conference on Human-Robot Interaction (HRI)* (pp. 201–208). Amsterdam, Netherlands.
- Triesch, J., Teuscher, C., Deák, G. O., & Carlson, E. (2006). Gaze following: Why (not) learn it? *Developmental Science*, 9(2), 125–147.
- Van Schendel, J. A., & Cuijpers, R. H. (2015, April). Turn-yielding cues in robot-human conversation. In K. D. M. Salem, A. Weiss, P. Baxter (Ed.), *4th International Symposium on New Frontiers in Human-Robot Interaction*. Canterbury, UK.
- Vázquez, M., Steinfeld, A., Hudson, S. E., & Forlizzi, J. (2014). Spatial and other social engagement cues in a child-robot interaction: Effects of a sidekick. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Bielefeld, Germany: ACM.
- Vertegaal, R., & Ding, Y. (2002). Explaining effects of eye gaze on mediated group conversations: Amount or synchronization? In *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)* (p. 41–48). New York, NY: ACM.
- Vertegaal, R., Slagter, R., Veer, G. van der, & Nijholt, A. (2001). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (SIGCHI)* (Vol. 3, p. 301–308). Seattle, WA.
- Vijayakumar, S., Conradt, J., Shibata, T., & Schaal, S. (2001, October). Overt visual attention for a humanoid robot. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Vol. 4, p. 2332–2337). Maui, HI.
- Wainer, J., Feil-Seifer, D., Shell, D. A., & Matarić, M. J. (2007, August). Embodiment and human-robot interaction: A task-based perspective. In *IEEE Proceedings of the International Workshop on Robot and Human Interactive Communication (RO-MAN)* (p. 872–877). Jeju Island, South Korea.
- Wang, N., & Gratch, J. (2010, April). Don't just stare at me! In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)* (p. 1241–1249). Atlanta, GA: ACM.
- Xu, T., Zhang, H., & Yu, C. (2013). Cooperative gazing behaviors in human multi-robot interaction. *Interaction Studies*, 14(3), 390–418.
- Yamazaki, Yamazaki, K., Burdelski, M., Kuno, Y., & Fukushima, M. (2010). Coordination of verbal and

- non-verbal actions in human-robot interaction at museums and exhibitions. *Journal of Pragmatics*, 42, 2398–2414.
- Yamazaki, K., Kawashima, M., Kuno, Y., Akiya, N., Burdelski, M., Yamazaki, A., et al. (2007, September). Prior-to-request and request behaviors within elderly day care: Implications for developing service robots for use in multiparty settings. In L. Bannon, I. Wagner, C. Gutwin, R. Harper, & K. Schmidt (Eds.), *ECSCW* (p. 61-78). Limerick, Ireland: Springer.
- Yonezawa, T., Yamazoe, H., Utsumi, A., & Abe, S. (2007, November). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In *Proceedings of the International Conference on Multimodal Interfaces (ICMI)* (p. 140-145). Nagoya, Japan: ACM Press.
- Yoshikawa, Y., & Shinozawa, K. (2006, Aug). Responsive Robot Gaze to Interaction Partner. In *Robotics: Science and Systems*. Philadelphia, PA.
- Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., & Miyamoto, T. (2006, October). The effects of responsive eye movement and blinking behavior in a communication robot. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4564–4569).
- Yu, C., Schermerhorn, P., & Scheutz, M. (2012, January). Adaptive eye gaze patterns in interactions with human and artificial agents. *ACM Transactions on Interactive Intelligent Systems*, 1(2).
- Zaraki, A., Mazzei, D., Giuliani, M., & De Rossi, D. (2014, April). Designing and evaluating a social gaze-control system for a humanoid robot. *IEEE Transactions on Human-Machine Systems*, 44(2), 157–168.
- Zheng, M., Moon, A., Croft, E. A., & Meng, M. Q.-H. (2015). Impacts of robot head gaze on robot-to-human handovers. *International Journal of Social Robotics*.

---

Henny Admoni is currently at Carnegie Mellon University. Authors' names and contact information: Henny Admoni, Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213 USA. Email: henny@cmu.edu. Brian Scassellati, Department of Computer Science, Yale University, New Haven CT 06511 USA. Email: scasz@cs.yale.edu.