Robot Behavior Toolkit:
Generating Effective Social Behaviors for Robots

Chien-Ming Huang, Bilge Mutlu
Department of Computer Sciences, University of Wisconsin–Madison
1210 West Dayton Street, Madison, WI 53706, USA
{cmhuang, bilge}@cs.wisc.edu

ABSTRACT

Social interaction involves a large number of patterned behaviors that people employ to achieve particular communicative goals. To achieve fluent and effective humanlike communication, robots must seamlessly integrate the necessary social behaviors for a given interaction context. However, very little is known about how robots might be equipped with a collection of such behaviors and how they might employ these behaviors in social interaction. In this paper, we propose a framework that guides the generation of social behavior for humanlike robots by systematically using specifications of social behavior from the social-sciences and contextualizing these specifications in an Activity-Theory-based interaction model. We present the *Robot Behavior Toolkit*, an open-source implementation of this framework as a Robot Operating System (ROS) module and a community-based repository for behavioral specifications, and an evaluation of the effectiveness of the Toolkit in using these specifications to generate social behavior in a human-robot interaction study, focusing particularly on gaze behavior. The results show that specifications from this knowledge base enabled the Toolkit to achieve positive social, cognitive, and task outcomes, such as improved information recall, collaborative work, and perceptions of the robot.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems – *human factors, software psychology*; H.5.2 [**Information Interfaces and Presentation**]: User Interfaces – *evaluation/methodology, user-centered design*; I.2.0 [**Artificial Intelligence**]: General – *Cognitive Simulation*

General Terms

Design, Human Factors

Keywords

Human-robot interaction, generating social behavior, Activity Theory, Robot Behavior Toolkit

HRI '12 Boston, Massachusetts USA
Copyright 2012 ACM 978-1-4503-1063-5/12/03 ...$10.00.

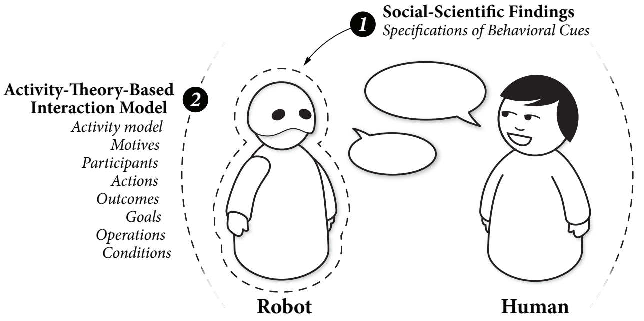
1
Social-Scientific Findings
Specifications of Behavioral Cues
Activity-Theory-Based Interaction Model
2
Activity model
Motives
Participants
Actions
Outcomes
Goals
Operations
Conditions
Robot
Human

Figure 1: Our framework that integrates (1) social-scientific specifications of human social behavior and (2) an interaction model inspired by Activity Theory to guide the generation of humanlike behavior for robots.

1. INTRODUCTION

Participants in human interactions routinely coordinate a set of social behaviors that support joint activities toward achieving social, cognitive, and task outcomes. For instance, teachers and students in the classroom employ social behaviors that support their teaching and learning goals. Public speakers display social behaviors that support their goals of effective communication, attention, and persuasion. Surgeons and their teams in operation rooms use social behaviors to communicate and coordinate their activities to improve the effectiveness of their work. Robots—as effective teachers, storytellers, and collaborators—must also employ a rich set of social behaviors to support their users' activities and goals.

Several studies in human-robot interaction, including our own work (e.g., [19, 21]), have modeled key social behaviors, implemented these behaviors on robots, and evaluated their effectiveness in supporting human activity (e.g., [4, 13, 25, 32]). These studies demonstrate the necessity of robots displaying appropriate social behaviors to effectively achieve social, cognitive, and task outcomes in human-robot interaction. In this paper, we seek further the study of appropriate social behavior for robots and explore the following research questions: How do we generate such behaviors systematically? How can we ensure that these generated behaviors support human activity? Can such behaviors reliably generate the intended social, cognitive, and task outcomes?

To answer these questions, we propose a framework that contextualizes an expandable knowledge base of social-scientific findings on human social behavior in an interaction model that facilitates human interaction based on Activity Theory [22] (see Figure 1). In this paper, we present an implementation of this framework, the *Robot Behavior Toolkit*[1], and an evaluation of the effectiveness of

[1] The open-source toolkit and its documentation can be accessed at http://hci.cs.wisc.edu/projects/rbt.

the social behaviors it generates for humanlike robots. The Toolkit offers a community-based repository for social-scientific findings, open to HRI researchers to use and to build on, and an open-source Robot Operating System (ROS) [24] module that integrates the behavioral specifications provided by these findings into an interaction model that supports human activity. The evaluation validates that the behaviors generated by our Toolkit based on a small, experimental set of behavioral specifications indeed creates improved social, cognitive, and task outcomes as predicted by literature on human-human interaction.

This work provides the following contributions. First, we present a system that offers a *community knowledge base* of social-scientific specifications for generating social behavior and an *open-source ROS module* for an interaction model that contextualizes behavioral specifications in human activity. Second, our evaluation, which focuses on gaze behavior, validates the effectiveness of the system in generating social behaviors that evoke key social, cognitive, and task outcomes. The evaluation also creates new knowledge on the specific outcomes that robots can generate using the subset of behavioral specifications explored in this work.

The next section provides background on approaches to generating robot behavior, on models of human activity and social interaction, and on related work from human-robot interaction on generating humanlike behavior in robots.

2. BACKGROUND

Approaches to Designing Robot Behaviors

Researchers and designers have developed and followed several approaches to designing interactive behaviors for embodied agents and robots, most notable of which follow principles of drama and film animatronics, animation techniques, and models of human social behavior. An example of the first approach is a puppeteering system that Hoffman et al. developed based on acting theory to support robotic live stage performers [11]. Similarly, Bruce et al. used lessons from dramatic acting to create believable behaviors for robots [6]. Researchers have also explored how animation principles might guide the design of interactive behaviors for humanlike robots [28] and robotic characters [31] and found that robots that employ principles such as anticipation and follow-through [29] are perceived to be more readable, appealing, approachable, and capable [28].

An example of using models of human behavior to design robot behavior is a software architecture developed by Breazeal and Scassellati [5], which generated facial expressions and behavioral responses to external stimuli that resembled infant-caregiver interaction. Another example is the BEAT system, which it generates appropriate gaze behaviors, gestures, and facial and prosodic expressions for animated agents based on models of human behavior [7]. Our previous work also explored how models of human gaze might guide the design of conversational gaze mechanisms for humanlike robots and showed that these mechanisms enable robots to establish appropriate conversational roles and rapport with their human partners [20, 21]. Finally, Holroyd et al. developed a Robot Operating System (ROS) module that generates a set of nonverbal behaviors to support engagement between a human and a robot based on observational studies of human engagement [13].

While all of these approaches to designing robot behavior have merit, we believe that human social behavior might serve as a rich resource for specifying robot behavior and a gold standard against which designs can be compared. Therefore, our approach draws on a collection of specifications of human behavior from research on human-human interaction to systematically generate robot behavior and to assess the extent to which these specifications enable robots to achieve communicative goals that humans achieve in social interaction.

Interaction Models

Research in psychology and human-computer interaction have proposed several paradigms to model human-human interaction and activities. The approaches that are most relevant to our work are Activity Theory, situated action models, and distributed cognition. *Activity Theory*, which serves as the basis for our work, offers a theoretical framework to understand human activity as a complex, socially situated phenomena and a set of principles that guide this understanding (see Leont'ev [17] and Nardi [23] for reviews of Activity Theory). Here we provide brief descriptions of each principle and how they inform the design of the interaction framework for our toolkit. More detail on the design of our system is provided in the next section.

The first principle, consciousness unifies attention, intention, memory, reasoning, and speech [30] toward understanding an activity. This principle informs the overall design of our toolkit; the concepts of attention and intention are captured in the *context model* and the *activity model*. Speech is considered an inseparable behavioral channel from and is synchronized with other behavioral channels. *Working memory* and *long-term memory* are used in the Toolkit to facilitate cognitive processes. The second principle is object-orientedness, which specifies that objects around which the activity is centered are “shared for manipulation and transformation by the participants of the activity” [15]. We represent this concept as *motives* in our *activity model*. The third principle of Activity Theory is the hierarchical structure of activity, which organizes activity into three levels: activity, action, and operation. Each level corresponds to a motive, goal, and conditions, respectively. An activity consists of a series of actions that share the same motive. Each action has a defined goal and a chain of operations that are regular routines performed under a set of conditions provided by the environment and the actions. We represent this hierarchy in our *activity model*.

The fourth principle of Activity Theory is internalization and externalization. Internalization is the process of transforming external actions or perceptions into mental processes, whereas externalization is the process of manifesting mental processes in external actions. Internalization and externalization are analogies to the processes of forming a *context model* and of generating output behavior, respectively. The fifth principle of Activity Theory is mediation. Activities are mediated by several external and internal tools such as physical artifacts that might be used in an activity and cultural knowledge or social experience that individual might have acquired. For instance, the *knowledge base* of behavioral specifications in our Toolkit serves as an internal tool that mediates activities between humans and robots.

Situated action models and *distributed cognition* propose alternative accounts of human activity. Situated action models posit that the nature of human activity and interaction is improvisatory and contingent and do not highlight concepts of motivation, goal, and consciousness in human activity [16]. Distributed cognition emphasizes the coordination among humans and artifacts toward achieving an overall system goal [12] and considers humans and artifacts as conceptually equivalent, both as agents in a cognitive system. This treatment challenges the role of artifacts as mediators in shaping human activity and behavior.

Among these frameworks, we find Activity Theory to offer the richest representation for interaction between humans and robots, because (1) its emphasis on mediation allows us to focus on the notion of context in human-robot interaction, (2) it decomposes activities into manageable layers for a robot to interpret and execute,

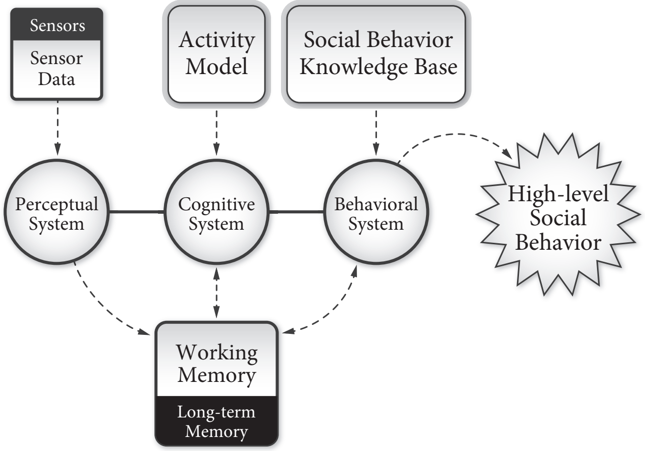

Figure 2: The Robot Behavior Toolkit consists of three subsystems—the perceptual, cognitive, and behavior systems; two memories—the working memory and long term memory; and supporting components—the activity model and knowledge base.

and (3) the notion of object-orientedness offers an essential representation for human-robot collaborative work. Therefore, we adopt Activity Theory as the theoretical basis for the design of our toolkit. A detailed comparison of the three approaches discussed here can be found in Nardi [23].

In this paper, we seek to couple two powerful concepts—using specifications of human social behavior as a resource for generating robot behavior and drawing on principles of Activity Theory to construct an activity model for human-robot interaction—toward achieving robots that systematically use social behavior to achieve positive social, cognitive, and task outcomes in human-robot interaction.

3. ROBOT BEHAVIOR TOOLKIT

Robot Behavior Toolkit is an implementation of our proposed framework—using social-scientific findings contextualized in an Activity-Theory-based interaction model to guide the generation of humanlike behavior for robots. The Toolkit is developed to be extensible and flexible to be used with a wide range of sensor devices and robotic platforms. To this end, we integrated the Toolkit with the ROS infrastructure. The Toolkit is written in Python and adopts XML as the main format for information sharing and storage within the Toolkit.

The Toolkit consists of three main subsystems and two memories (see Figure 2 for an illustration of the system architecture). The *perceptual system* takes sensor data as input and preprocesses it to an internal structure for later processing. The *cognitive system* uses the external information from the perceptual system and internal information from the *activity model* to form a context model of the current situation. The *behavior system* uses the context model to guide behavior formation based on behavioral specifications from the *knowledge base*. The output of the Toolkit is high-level behaviors defined in an XML format for the robot to execute. The following paragraphs provide more detailed descriptions of these system components. More detail on the Toolkit and its implementation is available on our project website: http://hci.cs.wisc.edu/projects/rbt.

Perceptual System

The *perceptual system* fuses sensor data from various sensor devices and transforms it into an internal data structure for the other toolkit components to use. In the current design, the perceptual system requires sensor data in an XML format. Each kind of sensor data (e.g., vision, audio, etc.) has its own XML specification (see our project page for details). This requirement helps standardize the information sharing between various sensor devices and our toolkit. Our design allows different sensor devices to send data to the perceptual system at different rates.

Memories

Informed by psychology research on memory [2, 3], two types of memory are used in our toolkit. The *working memory* stores data that is currently of interest to the subsystems. Conceptually, the working memory is a place where a subsystem stores information that other subsystems need for processing. For example, the cognitive system might store a context model of the current situation for the behavior system to access in order to generate the appropriate behavior output for the situation. If a concept or fact is repeatedly reinforced in the working memory, it is stored as 'knowledge' in the *long-term memory*. Information stored in the long-term memory has the potential to influence the cognitive system in context formation. This setup is analogous to the concept of mediation of internal tools (e.g., cultural knowledge and social experience) in Activity Theory.

Activity Model

An *activity model*, structured in XML format, specifies the activity that a robot wants to initiate. Figure 3 provides the representation for an example activity model. For each activity, a *motive* governs actions. Each action, by achieving its corresponding goal, helps to fulfill the motive of the activity. Each action may have several operations that are constrained by a set of conditions and that can be executed only when all the conditions are met. Actions have predefined *outcomes* such as 'task' and 'rapport' that are used in the process of coordinating generated behaviors (see *Behavior System*). Outcomes specify the orientation of an action. For instance, a *task* outcome indicates that the action is task-oriented.

In the example in Figure 3, the robot has a motive to clear the objects on a table. To fulfill the motive, the robot performs an action with the goal of removing the objects and, in this case, instructs the user to clear the table by moving the objects to boxes. The robot needs to ensure that 'user,' 'object,' and 'box' are present before performing the operation (i.e., giving the instruction).

Cognitive System

The *cognitive system* takes external and internal information and generates a set of triggers that form a context model of the current situation. This system consists of three main components: the *world manager*, *activity manager*, and *context generator*. The world manager keeps track of current environmental information (e.g., what objects are in the field of view of the robot and where

```
<Activity id=`1`>
  <Motive>clear(table)</Motive>
  <Description>Clear objects on table</Description>
  <Participants>Self, User1</Participants>
  <Action id=`1`>
    <Outcome>Task</Outcome>
    <Goal>disappear(object)</Goal>
    <Description>
      Instruct User1 to categorize object
    </Description>
    <Operation type=`utterance`>
      <Condition>present(User1)</Condition>
      <Condition>
        known(the blue object with two pegs)
      </Condition>
      <Condition> known(the blue box)</Condition>
      <Info turn=`end`>
        Could you help me put the blue object with
        two pegs into the blue box, please?
      </Info>
    </Operation>
    ...
```

Figure 3: An example activity model in which the robot instructs a human partner to clear objects on a table.

```
<rules>
  <rule id=`1`>
    <editor>editor_1</editor>
    <edit_time>timestamp_1</edit_time>
    <references>reference_id</references>
    <description>
      The referential gaze typically precedes the onset of corresponding
      linguistic reference by approimate 800 msec to 1000 msec.
    </description>
    <behavior_type>gaze</behavior_type>
    <trigger>linguistic_reference</trigger>
    <behavior>
      precede(toward(gaze, artifact), linguistic_reference, rand(800,
      1000))
    </behavior>
    <outcomes>task</outcomes>
  </rule>
  ...
```

Figure 4: An example behavioral specification on synchronizing referential cues in speech and gaze.

they are) using data from the perceptual system. This management provides a high-level abstraction for all processes in the Toolkit. The *behavior system* operates at an abstract level and does not have concrete details such as where an object might be located. Behavior realizers, processes that interpret behaviors generated by the Toolkit (see *Integration with ROS*), can query specific locations from the world manager when it is ready to execute the behaviors. Another advantage of this abstraction is providing accurate information about the external world, such as the most up-to-date information on the location of an object of interest that might be moved during processing.

The activity manager controls the flow of self-initiated activity specified in the *activity model* and examines whether the conditions are met for an operation to be executed and whether a goal is met so that it can proceed to the next action. Actions are organized in a queue structure in the order specified by the activity model. However, it is also possible to insert a new action in the queue when an unexpected event occurs, such as responding to spontaneous user input. If all operations under an action are executed, but the goal is not met, all operations will be re-executed. When the action queue is empty, the activity manager seeks to verify whether or not the motive of the activity is fulfilled.

The context generator uses internal (i.e., from the activity manager) and external (i.e., from the world manager) information to derive corresponding internal and external triggers, which together form the context model of the current situation. This model is represented in XML format and later used to guide the generation of robot behavior.

Knowledge Base

The *knowledge base* stores a collection of behavioral specifications from the social-scientific literature on human social behavior. The main purpose of using such a repository is to collectively organize relevant findings and to systematically apply these findings toward generating robot behaviors. Behavioral specifications are defined in XML format, as shown in Figure 4. Each specification has a *references* tag that can associate the specification with the source from which the specification is derived. Each specification may have multiple *triggers* that activate the specified behaviors. Conceptually, the context model from the cognitive system provides triggers used to retrieve behavioral specifications to generate robot behaviors. *behaviors* are described in a machine-readable function format. In the example illustrated in Figure 4, three arguments are provided for the *precede* function. The first argument precedes the second argument and the third argument tells the system how far ahead the first argument should precede the second argument. In this case, the action of gazing toward the artifact precedes corresponding linguistic reference by a period of time between 800 and 1000 milliseconds. This specification is derived from findings from research on human gaze, which suggest that referential gaze precedes corresponding linguistic reference by approximately 800 to 1000 milliseconds [10, 18]. Each specification also has *outcomes* that indicate the primary outcome of the specified behavior.

Behavior System

The *behavior system* generates humanlike behaviors based on the current context and behavioral specifications and includes three components: *behavior selector*, *behavior coordinator*, and *behavior generator*. The behavior selector uses triggers defined in the context model from the cognitive system to query the knowledge base for appropriate behavioral specifications. The behavior coordinator resolves conflicts and/or overlaps among specifications by prioritizing them. While this prioritization can be done in many ways, our current implementation uses *outcomes* as the criterion for prioritizing specifications. For instance, if the current action is task-oriented, rules with task outcomes are preferred over other outcomes, such as building rapport. Another main function of the behavior coordinator is to coordinate different behavioral channels (e.g., gaze, gesture, and nodding).

Finally, the behavior generator organizes coordinated behaviors in am XML format for execution. An example output behavior is shown in Figure 5. In this example, the robot uses two behavioral channels (i.e., gaze and speech) in synchrony. The timestamps specified for gaze behavior correspond to the speech timeline. Time periods for which no behaviors are specified are marked as *unspecified*. How unspecified periods might be interpreted depends on the particular developer's design decisions. These decisions are expressed in the behavior realizer, which is not a part of the Toolkit. Our current implementation uses an event-trigger mechanism, which directs the robot to continue the previous specified behavior until a new behavior is specified.

Integration with ROS

One of the goals of this work is to provide the HRI community with an open-source tool that generates humanlike behavior for a wide range of robot platforms. To this end, we integrated our Toolkit with the Robot Operating System (ROS) by implementing the Toolkit as a ROS node. ROS is becoming increasingly popular in the robotics and HRI communities and finding use in HRI research (e.g., [13, 25]). Figure 6 illustrates the current integration

```
<behaviors>
  <channel type=`gaze`>
    <action endTime=`214.5` startTime=`0` target=`unspecified`/>
    <action endTime=`1160` startTime=`214.5` target=`the green
      object with one peg`/>
    <actoin endTime=`2735.4` startTime=`1160` target=`unspecified`/>
    <action endTime=`3597` startTime=`2735.4` target=`the red box`/>
    <action endTime=`4308` startTime=`3597` target=`unspecified`/>
    <action endTime=`4963` startTime=`4308` target=`listener`/>
  </channel>
  <channel type=`speech`>
    Could you help me put the green object with one peg into the red
    box, please?
  </channel>
</behaviors>
```

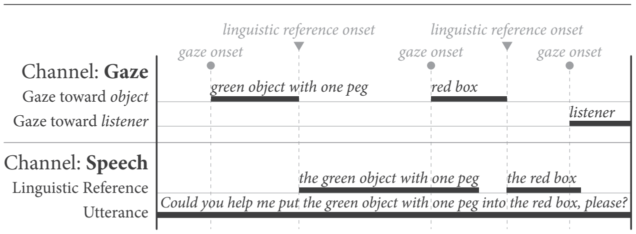
linguistic reference onset
linguistic reference onset
gaze onset
gaze onset
gaze onset
Channel: Gaze
Gaze toward object
Gaze toward listener
green object with one peg
red box
listener
Channel: Speech
Linguistic Reference
Utterance
the green object with one peg
the red box
Could you help me put the green object with one peg into the red box, please?

Figure 5: An example behavior output generated by the Toolkit in XML (top) and in visual representation (bottom).

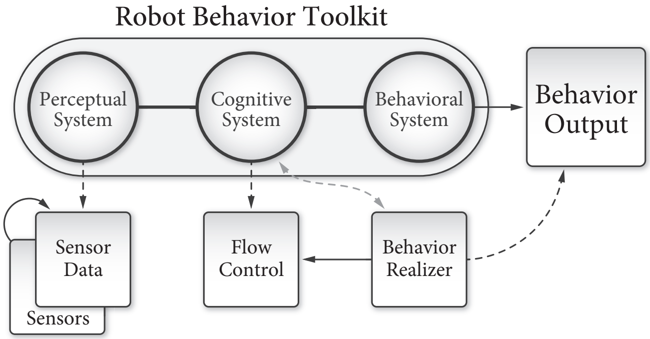

Figure 6: Information flow between our Toolkit and ROS. Rounded squares represent *topics*. Solid and dashed lines denote *publishing* and *subscribing* to topics, respectively. Light dashed lines denote *service* communication between nodes.

of the Toolkit into ROS. In our future work, we plan to improve this integration by migrating each subsystem into a ROS node and using ROS communication protocols to establish information flow among the subsystems.

Various sensors available in ROS can be used by publishing gathered sensor data for the Toolkit to use. Similarly, robotic platforms available in ROS can work with the Toolkit by subscribing to the *behavior output*. *Behavior realizer* is a ROS node that interprets the behavior output generated by the Toolkit and sends the commands necessary to execute the behavior to the robot. The behavior realizer uses the *service* communication with the world manager to obtain information on the environment. The behavior realizer also informs the Toolkit of the status of the execution of specified behaviors (e.g., the completion of an action) through the *flow control*. A dedicated behavior realizer is needed for each robot platform to realize the behavioral output. This modularity makes it possible for the Toolkit to work with a wide range of robot platforms. By the time of the writing of this paper, we have integrated and tested the Toolkit with two robot platforms, Wakamaru and PR2 (simulated in Gazebo). The study reported in the next section uses the integrated system of the Toolkit, ROS, and the Wakamaru robot.

4. EVALUATION

Many aspects of the Toolkit need to be evaluated. One such aspect is its *usability*: Is the Toolkit easy to use by the developers? Another aspect is *effectiveness*: Does the Toolkit generate robot behaviors that are effective in achieving communicative goals? While both evaluations are valuable and necessary in assessing our Toolkit, the current paper focuses on the evaluation of the effectiveness of the Toolkit in generating humanlike social behavior for robots, focusing particularly on *gaze* behavior.

4.1 Hypotheses

Drawing on gaze behavior literature, we have developed three hypotheses on how the gaze behaviors generated by our toolkit might affect social, cognitive, and task outcomes in human-robot interaction against different baseline specifications.

Hypothesis 1: Participants will recall the information that the robot presents to them more accurately when the robot employs the gaze behaviors generated by our toolkit than they will when the robot employs alternative behaviors. The basis of this hypothesis is the finding from gaze literature that gaze cues clarify what is being referred to in speech and improve story comprehension [26].

Hypothesis 2: Participants' performance in a collaborative sorting task will be higher when the robot employs gaze behaviors generated by our toolkit than it will when the robot employs alternative behaviors. This hypothesis builds on prior work in human-robot interaction that suggests that appropriately timed gaze cues of a robot facilitate the effective locating of information among distractions [27].

Hypothesis 3: Participants will evaluate the robot as more natural, likable, and competent when it employs gaze cues generated by our toolkit than they will when the robot generates alternative behaviors. This prediction follows findings from our prior research that gaze cues shape the favorability of the robot [21, 19].

4.2 Participants

A total of 32 participants were recruited for the evaluation study. All participants were native English speakers from the Madison, Wisconsin area with an average age of 24.9 years, ranging between 18 and 61. Average familiarity with robots among the participants was relatively low (M=3.25, SD=1.67) and verage familiarity with the experimental tasks was also low (M=2.13, SD=1.21) when measured by seven-point rating scales.

4.3 Experimental Design, Task, & Procedure

We tested our hypotheses in a laboratory experiment, which involved two human-robot interaction scenarios in order to increase the generalizability of our findings across task contexts. In the first scenario, the robot told participants the story of the 12 signs of the Chinese Zodiac (see top picture in Figure 7). In its story, the robot referred to a set of cards that were laid on a table located between the robot and the participant. The cards showed pictures of the 12 animal characters and the figure mentioned in the story. The second scenario involved a collaborative categorization task (see bottom picture in Figure 7). In the task, the robot instructed the participants to categorize a set of colored lego blocks into different colored boxes. There were 15 blocks with different colors, sizes, and heights and two colored boxes laid on the table located between the robot and the participant. The participant did not know how the each block should be categorized and had to wait for instructions from the robot to place each block into its respective box. We used a pre-recorded human voice for the robot's speech and modulated its pitch to create a gender-neutral voice.

We manipulated the specifications in the knowledge base of our toolkit to create the following four conditions for both tasks:

(1) **Humanlike**: The robot exhibited gaze behaviors generated by our toolkit using the following social behavioral rules (full specifications of these rules can be found on the project website):

- ∗ Referential gaze precedes linguistic reference by approximately 800 to 1000 milliseconds [10, 18].
- ∗ The speaker looks toward the listener at the end of a turn [8].
- ∗ The speaker looks toward the person whom he/she is greeting [14].

(1) **Delayed**: The robot showed the same behaviors as it did in the *humanlike* condition except that the behaviors were delayed, resembling the timings of the listener as opposed to that of the speaker, e.g., referential gaze following the onset of the linguistic reference by approximately 500 to 1000 milliseconds [9].

(3) **Incongruent**: The robot followed the timings in the *humanlike* condition, but looked toward an object that was different from what was referred to in the linguistic reference.

(4) **No-gaze**: The robot did not display any gaze behaviors other than tracking the participant's face.

In all conditions, the robot tracked the participant's face when the specified gaze behavior involved looking toward the listener. The linguistic references in the robot's speech were manually marked.

The study followed a between-participants design. Participants were randomly assigned to one of the four conditions. There were

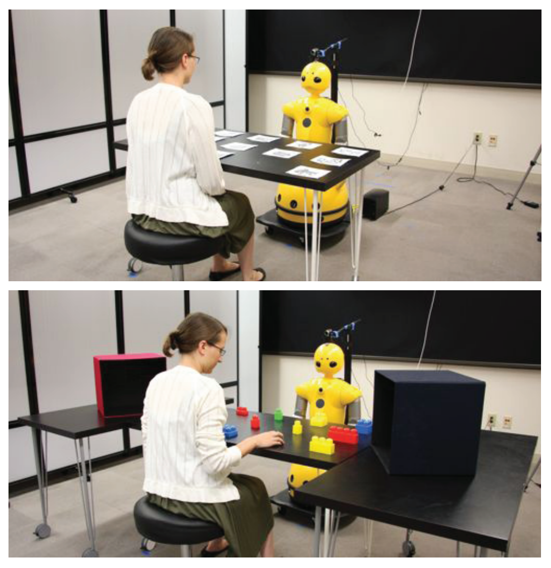

Figure 7: The setup of the storytelling (top) and collaborative work (bottom) tasks in the experiment.

four male and four female participants in each condition. The first and second task involved a total of one and eight trials, respectively. In each trial of the second task, the order in which the robot referred to the objects was randomized. At the beginning of the study, the experimenter provided the participants with a brief introduction of the goals of the study and obtained informed consent. After the first task, the participants took a three-minute break while the experimenter prepared for the second task. After completing the second task, the participants were asked to complete a recall test about the story. They then filled out a post-experiment questionnaire. At the end of the study, the experimenter interviewed and debriefed the participants. The experiment took approximately 30 minutes. The participants received $5 for their participation.

4.4 Measurement

The two independent variables in our study were the manipulation in the behavioral specifications for the robot's gaze behavior and participant gender. We measured two types of dependent variables, objective and subjective.

Objective: Following the storytelling task, we measured the participants' recall of the details of the robot's story. A total of 10 questions were asked in the recall test. All questions were related to the order in which the characters were presented in the story about the signs of the Chinese Zodiac. The questions followed true-or-false, multiple-choice, or multi-select formats. An example question is provided below.

> Q: The Dragon is before the Rabbit in the Zodiac cycle.
> A: "False"

Following the collaborative work task, we measured the time that the participant took to locate objects to which the robot referred. In particular, we measured the time between the end of the linguistic reference and one of the following cases: (1) the participant's last gaze toward the object before moving the object, (2) the participant touching the object, or (3) the participant reaching for the object. This measure served as a measure of task performance and captured how fast the participants located the information needed to complete the task.

Subjective: We used a post-experiment questionnaire to measure participants' perceptions of the robot in terms of naturalness of behavior, likability, and competence. The questionnaire also included several manipulation-check questions. Seven-point rating scales were used in all questionnaire items. The naturalness scale, which consisted of seven items, measured how participants perceived the naturalness of the robot's behavior. The likability scale consisted of 10 items, which measured participants' ratings of the likability of the robot. The competence scale, which consisted of 14 items, measured the participants' perceptions of the robot's competence in the task and its overall competence. Item reliabilities for naturalness (Cronbach's $\alpha = .79$), likability (Cronbach's $\alpha = .90$), and competence (Cronbach's $\alpha = .85$) scales were sufficiently high.

4.5 Results

We used one-way analyses of variance (ANOVA) to analyze data from our manipulation checks and two-way analyses of variance for objective and subjective measures.

Manipulation checks: To test whether the manipulation in the robot's gaze behavior was successful, we asked participants whether the robot's gaze seemed to be random, whether the timing of when the robot looked toward objects seemed right, whether the timing of when the robot referred to an object and looked toward it matched, and whether the robot's gaze and speech were synchronized. The results showed that the participants were able to identify the differences across conditions in the majority of these measures; the gaze manipulation had a significant effect on whether the participants found the robot's gaze to be random, $F(3, 28) = 3.55, p = .027$, $\eta_p^2 = 0.275$, whether the timing of when the robot looked toward objects seemed right, $F(3, 28) = 11.83, p < .001, \eta_p^2 = 0.559$, whether they thought that the timing of when the robot referred to an object and looked toward it matched, $F(3, 28) = 33.42$, $p < .001, \eta_p^2 = 0.782$, and whether they found the robot's gaze and speech to be synchronized, $F(3, 28) = 4.96, p = .007, \eta_p^2 = 0.347$. There was no main effect of the manipulation on whether the participants thought that the robot looked toward them at the right time, $F(3, 28) = 0.78, p = .514, \eta_p^2 = 0.077$. An explanation for this result is that the robot looked toward the participants for the majority of the time by tracking their faces including the *no gaze* condition.

Objective: Our first hypothesis predicted that participants would have better recall of the story told by the robot when it displayed humanlike gaze behavior than they would when the robot displayed alternative behaviors. Our data confirmed this hypothesis. The number of correct answers out of ten questions in the recall test were on average 7.38 ($SD = 2.67$), 4.25 ($SD = 2.49$), 4.50 ($SD = 1.41$), and 4.75 ($SD = 1.39$) for humanlike, delayed, incongruent, and no gaze, respectively. The analysis of variance found a significant main effect of the robot's gaze behavior on recall accuracy, $F(3, 24) = 4.51, p = .012, \eta_p^2 = 0.360$. Pairwise comparisons using Tukey's HSD test revealed that the recall performance of the participants in the humanlike condition significantly outperformed those of the participants in delayed, $F(1, 24) = 10.45, p = .004, \eta_p^2 = 0.303$, incongruent, $F(1, 24) = 8.84$, $p = .007, \eta_p^2 = 0.269$, and no-gaze, $F(1, 24) = 7.37, p = .012$, $\eta_p^2 = 0.235$, conditions. We also found a main effect of gender on participants' recall accuracy; male participants had better recall performance than female participants had, $F(1, 24) = 5.22$, $p = .031, \eta_p^2 = 0.179$. These results are illustrated in Figure 8. Post-hoc tests showed that male participants' recall was significantly better than that of female participants in the humanlike condition, $F(1, 24) = 5.65, p = .026, \eta_p^2 = 0.191$.

The second hypothesis predicted that the participants would show better task performance—measured by the time that participants took to locate objects that the robot referred to—in the col-

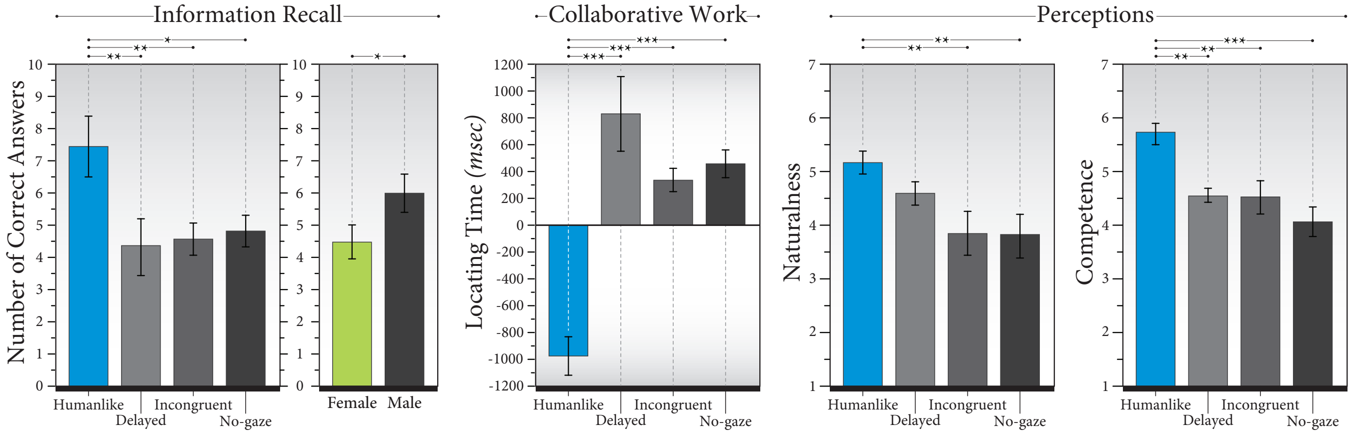

Figure 8: Results on information recall, collaborative work, and perceptions. (*), (**), and (***) denotes $p < .05$, $p < .01$, and, $p < .001$, respectively.

laborative work task when the robot displayed humanlike gaze behavior than they would when the robot showed alternative behaviors. This hypothesis was also supported by our data. When the end of the linguistic reference was represented by 0, the average time in milliseconds that the participants took to locate the object were -975.26 ($SD = 405.43$), 829.51 ($SD = 779.04$), 334.78 ($SD = 241.00$), and 457.05 ($SD = 292.51$) for humanlike, delayed, incongruent, and no-gaze conditions, respectively. The analysis of variance found a main effect of the robot's gaze behavior on locating time, $F(3, 24) = 23.22$, $p < .001$, $\eta_p^2 = 0.225$. Pairwise comparisons using Tukey's HSD test revealed that participants in the humanlike condition located the objects that the robot referred to in a significantly shorter time than participants in the delayed, $F(1, 24) = 61.12$, $p < .001$, $\eta_p^2 = 0.662$, incongruent, $F(1, 24) = 32.20$, $p < .001$, $\eta_p^2 = 0.578$, and no-gaze, $F(1, 24) = 38.50$, $p < .001$, $\eta_p^2 = 0.610$, conditions did (see Figure 8).

Subjective: The third hypothesis predicted that the participants would perceive the robot to be more natural, likable, and competent in the humanlike condition than they would in the other conditions. Our data provided partial support for this hypothesis. Results from the subjective measures showed a main effect of the gaze manipulation on participants' perceptions of the robot's naturalness, $F(3, 24) = 4.05$, $p = .018$, $\eta_p^2 = 0.336$, and competence, $F(3, 24) = 7.79$, $p < .001$, $\eta_p^2 = 0.493$, while the effect of the manipulation on measures of likability was not significant, $F(3, 24) = 1.46$, $p = .249$, $\eta_p^2 = 0.155$. In particular, participants in the humanlike condition rated the robot to be more natural than they did in the incongruent, $F(1, 24) = 8.44$, $p = .009$, $\eta_p^2 = 0.260$, and no gaze, $F(1, 24) = 8.67$, $p = .007$, $\eta_p^2 = 0.265$, conditions. Male participants in the incongruent condition rated the robot to be less natural than female participants did, $F(1, 24) = 4.69$, $p = .041$, $\eta_p^2 = 0.163$. However, participants in both the humanlike and delayed conditions found the robot to be equally natural, $F(1, 24) = 1.58$, $p = .221$, $\eta_p^2 = 0.062$. The participants in the humanlike condition rated the robot to be more competent than they did in the delayed condition, $F(1, 24) = 10.81$, $p = .003$, $\eta_p^2 = 0.311$, incongruent, $F(1, 24) = 11.14$, $p = .003$, $\eta_p^2 = 0.317$, and no-gaze, $F(1, 24) = 21.37$, $p < .001$, $\eta_p^2 = 0.471$, conditions. These results are also illustrated in Figure 8.

4.6 Discussion

The results provide support for the majority of our hypotheses in measures of information recall, collaborative work, and perceptions of the robot. Participants had better recall of information and located objects that the robot referred to faster when it used humanlike gaze behavior generated by our toolkit than they did when the robot displayed alternative behaviors. Moreover, participants found the robot to be more natural and competent when it exhibited humanlike gaze behavior than they did in other baseline conditions. These results suggest that the gaze behaviors that our toolkit generated were effective in evoking social, cognitive, and task outcomes in human-robot interaction, as predicted by our knowledge of human-human behavior. They also confirm that gaze cues serve as powerful communicative signals in storytelling and instructional settings.

Our data indicates that the participants in the delayed, incongruent, and no-gaze conditions needed roughly 300 to 800 milliseconds to locate the object that the robot referred to after it completed the linguistic reference to the object (see Figure 8). This result is consistent with findings in the gaze literature; in the absence of speaker gaze cues, partners look toward the object of reference approximately 200 to 300 milliseconds after they hear the reference [1] and approximately 500 to 1000 milliseconds after the onset of the spoken reference [9]. The result suggests that the participants in the baseline conditions (i.e., delayed, incongruent, and no-gaze) relied primarily on the robot's speech to locate the object of reference, while those in the humanlike gaze condition used gaze information to locate the object, completing the task even before the robot ended the linguistic reference.

5. GENERAL DISCUSSION

In this paper, we proposed a framework that uses behavioral specifications from the social sciences and an interaction model inspired by Activity Theory to guide the systematic generation of humanlike behavior for robots. We presented *Robot Behavior Toolkit*, an implementation of this framework, and evaluated its effectiveness in generating social behaviors that achieve positive social, cognitive, and task outcomes in a human-robot interaction study with two scenarios. Our findings highlight the potential of our toolkit for generating effective robot behaviors and confirm the findings from previous research including our own that, using humanlike social behavior effectively, robots can achieve significant social, cognitive, and task improvements in human-robot interaction.

This work also showed that a small number of behavioral specifications are sufficient to generate robot gaze behaviors that achieve significant social, cognitive, and task improvements in human-robot interaction. While our study used a small number of behavioral specifications for gaze behavior for experimental purposes, the Toolkit offers the potential to realize complex humanlike behaviors by combining a large number of specifications for multiple channels of behavior, which is a significant challenge when hard-coding behavioral specifications into robots. The Toolkit also offers social scientists and HRI researchers the ability to validate new behavioral specifications by realizing them in interactive human-robot interaction scenarios.

The Toolkit and our current evaluation, however, are not without limitations. Here we discuss these limitations and the future work that might address them. First, the evaluation of the Toolkit used simulated sensor data, as we focused on generating robot behavior rather than recognizing human behavior. However, more investigation is needed to understand how our Toolkit might function in more realistic interactive settings in which the recognition of human activity and the environment might be incomplete due to unreliable sensor data. Furthermore, how our system might support more interactive tasks, e.g., tasks that require significant input from human partners, remains an open question. Second, our evaluation focused on assessing the effectiveness of the Toolkit in generating humanlike robot behavior and whether the Toolkit is easy to use by developers is unknown. In our future work, we plan to conduct usability tests to assess usability and to gain a better understand of how the design of the Toolkit might be extended to address the needs of developers and HRI researchers.

Third, the current paper focuses on gaze behavior and a small number of behavioral specification in order to achieve a proof-of-concept evaluation of our system. As a next step, we plan to extend our repository of behavioral specifications to include a wider range of behavioral channels and to investigate interactions among these channels of behavior. Fourth, our current design of the Toolkit requires behavioral specifications to be unchanging and explicitly entered into a repository. However, machine learning techniques might allow robots to generate new specifications and modify existing specifications based on experience and we hope to explore this potentially fruitful area in our future work. Fourth, the Toolkit in its current design selects behavioral specifications based on the target outcomes specified in the activity model. We plan to explore alternative ranking mechanism for behavioral specifications, as we envision the Toolkit to include a large number of specifications, which might suggest conflicting behavioral outputs.

We hope that the *Robot Behavior Toolkit* serves as a useful resource for the HRI community and inspires further development in designing effective social behaviors for robots.

6. ACKNOWLEDGMENTS

National Science Foundation award 1017952 supported this work.

7. REFERENCES

[1] G. Altmann and Y. Kamide. Now you see it, now you don't: Mediating the mapping between language and the visual world. In *The interface of Language, Vision, and Action: Eye Movements and The Visual World*, pages 347–386. Psychology Press, 2004.

[2] R. C. Atkinson and R. M. Shiffrin. Human memory: A proposed system and its control processes. In *The psychology of learning and motivation: Advances in research and theory*, volume 2. Academic Press, 1968.

[3] A. D. Baddeley and G. Hitch. Working memory. In *The psychology of learning and motivation: Advances in research and theory (Vol. 8)*. Academic Press, 1974.

[4] C. Breazeal, C. Kidd, A. Thomaz, G. Hoffman, and M. Berlin. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *Proc IROS '05*, pages 708–713, 2005.

[5] C. Breazeal and B. Scassellati. How to build robots that make friends and influence people. In *Proc IROS '99*, pages 858–863, 1999.

[6] A. Bruce, J. Knight, S. Listopad, B. Magerko, and I. Nourbakhsh. Robot improv: using drama to create believable agents. In *Proc ICRA '00*, volume 4, pages 4002–4008, 2000.

[7] J. Cassell, H. H. Vilhjalmsson, and T. Bickmore. Beat: the behavior expression animation toolkit. In *Proc SIGGRAPH '01*, pages 477–486, 2001.

[8] S. Duncan. Some signals and rules for taking speaking turns in conversations. *J. Personality and Social Psychology*, 23(2):283–292, 1972.

[9] B. Fischer. Attention in saccades. In *Visual attention*, pages 289–305. Oxford University Press, New York, 1998.

[10] Z. M. Griffin. Gaze durations during speech reflect word selection and phonological encoding. *Cognition*, 82(1):B1–B14, 2001.

[11] G. Hoffman, R. Kubat, and C. Breazeal. A hybrid control system for puppeteering a live robotic stage actor. In *Proc RO-MAN '08*, pages 354–359, 2008.

[12] J. Hollan, E. Hutchins, and D. Kirsh. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM TOCHI*, 7:174–196, 2000.

[13] A. Holroyd, C. Rich, C. L. Sidner, and B. Ponsler. Generating connection events for human-robot collaboration. In *Proc RO-MAN '11*, pages 241–246, 2011.

[14] A. Kendon and A. Ferber. A description of some human greetings. *Comparative ecology and behavior of primates*, pages 591–668, 1973.

[15] K. Kuutti. Activity theory as a potential framework for human-computer interaction research. In *Context and Consciousness: Activity Theory and Human-Computer Interaction*. The MIT Press, 1996.

[16] J. Lave. *Cognition in Practice: Mind, mathematics, and culture in everyday life*. Cambridge University Press, 1988.

[17] A. Leont'ev. *Activity, Consciousness, and Personality*. Prentice-Hall, 1978.

[18] A. S. Meyer, A. M. Sleiderink, and W. J. M. Levelt. Viewing and naming objects: eye movements during noun phrase production. *Cognition*, 66(2):B25–B33, 1998.

[19] B. Mutlu, J. K. Hodgins, and J. Forlizzi. A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proc HUMANOIDS '06*, 2006.

[20] B. Mutlu, T. Kanda, J. Forlizzi, J. Hodgins, and H. Ishiguro. Conversational gaze mechanisms for humanlike robots. *ACM Transactions on Interactive Intelligent Systems*, 1, 2012.

[21] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, and N. Hagita. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proc HRI '09*, pages 61–68, 2009.

[22] B. A. Nardi. Activity theory and human-computer interaction. In *Studying Context: A Comparison of Activity Theory, Situated Action Models, and Distributed Cognition*. The MIT Press, 1996.

[23] B. A. Nardi. Activity theory and human-computer interaction. In *Context and Consciousness: Activity Theory and Human-Computer Interaction*. The MIT Press, 1996.

[24] M. Quigley, K. Conley, B. P. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng. Ros: an open-source robot operating system. In *ICRA Workshop on Open Source Software*, 2009.

[25] C. Rich, B. Ponsleur, A. Holroyd, and C. L. Sidner. Recognizing engagement in human-robot interaction. In *Proc HRI '10*, pages 375–382, 2010.

[26] D. C. Richardson and R. Dale. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive Science*, 29(6):1045–1060, 2005.

[27] M. Staudte and M. W. Crocker. Visual attention in spoken human-robot interaction. In *Proc HRI '09*, pages 77–84, 2009.

[28] L. Takayama, D. Dooley, and W. Ju. Expressing thought: Improving robot readability with animation principles. In *Proc HRI '11*, pages 69–76, 2011.

[29] F. Thomas and O. Johnston. *The Illusion of Life: Disney Animation*. Abbeville Press, 1981.

[30] L. S. Vygotsky. Consciousness as a problem in the psychology of behaviour. In *Collected Works: Questions of the Theory and History of Psychology*. Pedagogika, Moscow, 1925/1982.

[31] R. Wistort. Only robots on the inside. *Interactions*, 17:72–74, March 2010.

[32] A. Yamazaki, K. Yamazaki, Y. Kuno, M. Burdelski, M. Kawashima, and H. Kuzuoka. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proc CHI '08*, pages 131–140, 2008.