Believable
Robot Characters

Reid Simmons, Maxim Makatchev, Rachel Kirby, Min Kyung Lee,
Imran Fanaswala, Brett Browning, Jodi Forlizzi, Majd Sakr

■ *Believability of characters has been an objective in literature, theater, film, and animation. We argue that believable robot characters are important in human-robot interaction, as well. In particular, we contend that believable characters evoke users' social responses that, for some tasks, lead to more natural interactions and are associated with improved task performance. In a dialogue-capable robot, a key to such believability is the integration of a consistent story line, verbal and nonverbal behaviors, and sociocultural context. We describe our work in this area and present empirical results from three robot receptionist test beds that operate "in the wild."*

The way people interact with robots appears to be fundamentally different from how they interact with most other technologies. People tend to ascribe a level of intelligence and sociability to robots that influences their perceptions of how the interactions should proceed. If a robot, for its part, adheres to this ascription, then the interactions tend to be natural and engaging; if not, they can be discordant. Our goal is to develop autonomous robotic systems that can sustain such natural and engaging social interactions with untrained users.

Our approach is to develop believable robot characters. In this context, *believable* means an illusion of life. Bates (1994) writes, "There is a notion in the Arts of 'believable character.' It does not mean an honest or reliable character, but one that provides the illusion of life, and thus permits the audience's suspension of disbelief …. Traditional character animators are among those artists who have sought to create believable characters …." Think of animated characters such as the magic carpet in Disney's *Aladdin* or the teapot and candlestick in *Beauty and the Beast* — not people, in any sense, but engaging, lifelike characters.

Perhaps a more apt analogy, though, is to view robots as actors performing in a human environment (Hoffman 2011). As early as the 19th century, acting theories, such as Delsarte's method (Stebbins 1886), placed emphasis on external actions as a key to believability. For example, an actor's dialogue should be rich in verbal and nonverbal expression of emotions and personality traits. While modern acting methods (Stanislavski

2008) instead focus on cohesive representations of underlying developments in the character, these are still manifested in a performer's surface actions. Thus, both traditions of acting emphasize the need for both verbal and nonverbal richness and continuity with character.

These ideas from animation and drama provide inspiration for our approach to developing believable characters. In general, a believable character should be cognizant of the people with whom it is interacting and exhibit behaviors consistent with the social norms of such interactions. Believability of robots, in addition to increasing a user's enjoyment while interacting, helps to make the interaction more natural by increasing predictability — the robot acts more like people expect. In this article, we describe our approaches to developing believable characters, focusing on both verbal and nonverbal behaviors. In particular, we present our use of *dramatic structure* with rich backstory and evolving story line, verbal and nonverbal *social behaviors,* and believable *culturally specific* characters.

To test our approach, we have developed several robot characters that operate long term at the Carnegie Mellon campuses in both Pittsburgh, Pennsylvania, and Doha, Qatar. The robots are similar in that they each play the role of a robot receptionist ("roboceptionist"), providing directions and general information about the university and environs. They each feature a cartoonlike, but expressive, three-dimensional graphical head and sensors that allow them to track and respond to people's actions. They all receive typed input and respond using text-to-speech with lip-syncing. They differ mainly in the characters themselves — including both male and female, human and machine, American and Arab. The differences are exhibited through a combination of facial features, voice, language used, nonverbal behaviors, and backstory.

Most research in human-robot interaction has involved laboratory experiments, where people and robots interact under relatively controlled conditions. While this is a valuable source of data, it may not be indicative of the types of interactions that people will have with deployed systems. An alternate approach, which we have followed for a number of years, is to place robots "in the wild," where people interact with them if, and when, they choose. While it is typically more difficult to evaluate such robots, due to a lack of ground truth measurement of user states, we believe that the interactions are much more natural and better capture the range of interactions that people will exhibit in actual settings. For instance, in interactions with our roboceptionists, we see many instances of foul language, personal questions, and marriage proposals — the types of interactions one is less likely to observe in laboratory settings,

where people are conscious of being recorded. Having a robot publicly accessible also enables us to capture interactions with diverse user groups, such as support staff and visiting parents and children, that represent challenges to conventional routes of subject recruitment. In such uncontrolled settings, we typically need to log and analyze hundreds or thousands of interactions over weeks, or even months, to find significant results. To date, our approach has yielded insight into various aspects of interaction including what dialogue topics people typically choose, how social behaviors manifest themselves and affect the outcomes of interactions, how display of emotion and mood affect the way people interact, and how interactions differ across cultures.

The next section presents our robot test beds. We then describe our approach to designing believable characters, including how we incorporate dramatic structure and emotions into the dialogue, how we utilize verbal and nonverbal behaviors to increase believability, and how we develop believable ethnically specific characters. We then present an empirical account of how our design approach affects the experience of interacting with robot characters in the wild and finally provide conclusions and future work.

Robot Test Beds

To test our approach, we wanted to place robots in situations where people would encounter them often and have the option of deciding whether to interact with, or ignore, them. The hypothesis is that believable robot characters would prove to be engaging and would attract people to interact with, and exhibit social behaviors toward, the robots. To that end, we wanted the robots to have a role that would be familiar to people and one where they would have occasional need for the robots' assistance. Thus, we chose to develop robot receptionists that greet visitors, provide information, such as directions to people's offices, and talk about their "life."

The first roboceptionist, named Valerie, was deployed in November 2003 (Gockley et al. 2005). It is housed in a custom-built booth in a corridor near the main entrance of the Robotics Institute at Carnegie Mellon, Pittsburgh (figure 1). The robot has a 15-inch monitor that displays a graphical face, mounted on a Directed Perception pan-tilt head atop a (stationary) RWI B21r base. The robot perceives people using a SICK laser mounted underneath the ledge of the booth. People interact with the roboceptionist by typing to it, and it responds vocally using the Cepstral text-to-speech system.[1] An auxiliary monitor over the keyboard shows people what they are typing and displays the IRB notification. Visitors affiliated with the

university can also swipe their ID cards to identify themselves to the robot.

Tank (figure 2) replaced Valerie in October 2005. The hardware and most of the software used is the same as Valerie's, but we changed the face, voice, and the character's backstory and story line. In addition, the booth was decorated to reflect Tank's backstory (worked for NASA and the CIA).

In 2008, Hala was deployed in the main atrium of the Carnegie Mellon campus in Doha, Qatar. Unlike the Pittsburgh roboceptionists, Hala is situated at a counter next to human receptionists and security personnel (figure 3). Hala can accept input in both Arabic and English, and responds in whichever language the user types, using the Acapela speech synthesis system.

The three robots all display an expressive graphical head, based on AT&T's SimpleFace. To date, the heads have been fairly simple and "cartoon-like" (figure 4(a–c)) to avoid raising undue expectations about the robot's capabilities. For the next version of Hala, however, we are shifting toward increased visual realism and adding visual cues of ethnicity (figure 4c).

For small motions, the head moves within the screen, while for larger motions the whole screen moves on the pan-tilt head. Phoneme outputs from the text-to-speech systems are used to automatically generate and synchronize lip movements with the speech audio. In addition, a scripting language was implemented that enables developers to define facial expressions that can be synchronized with speech.

The roboceptionists interact with people primarily in a question-answering mode, waiting for typed input (prompting if none is forthcoming). While signage helps guide people in interacting, they are largely left to themselves to discover what the roboceptionist knows and can talk about. In fact, the roboceptionists are fairly limited — they can tell what offices people are in (using an online university database), give directions around campus, answer questions about the weather around the world (by requesting and parsing free weather web services), provide the current date and time, and answer some specific questions about their locales. For instance, Tank can talk about Carnegie Mellon, Pittsburgh and, of course, the Steelers. In addition, each robot has a fairly rich backstory and evolving story line that it can speak about, which are described in the following section.

To handle input utterances, we adapted the AINE system, an open-source chatbot.[2] AINE provides a language for designing template-based rules for parsing text. Some of the rules generate scripted output, while others invoke special procedures (such as the ones for looking up people's offices and handling weather requests).
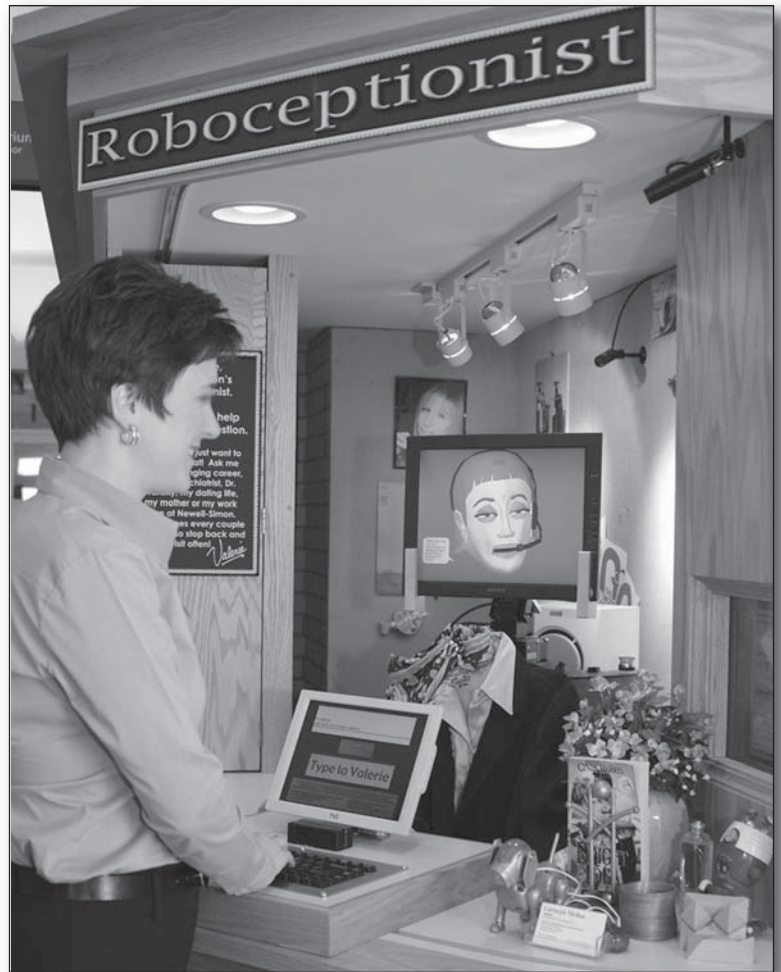
In addition to responding to users' utterances,

Figure 1. Valerie at Work in Her Booth.

the roboceptionists engage in some spontaneous behaviors. For instance, they greet people who are passing by, using a learned classifier to distinguish people who are likely to stop from those who are not, and focusing on the former. If people are standing nearby, the robot will encourage them to approach and interact. When no one is interacting, the roboceptionist will periodically engage in a simulated phone call, along the lines of what a human receptionist might do.

Designing Believable Robot Characters

Our primary approach to designing believable characters is to focus on the content and behavior of the character. The goal is to develop synthetic characters that are memorable and cohesive. Inspired by dramaturgy and acting theories, such as those of Delsarte (Stebbins 1886) and Stanislavski (2008), we strive to have the character's look, voice, personality, use of language, nonverbal

Figure 2. Tank Attending to a User.

expressions, backstory, and evolving story line be unified and contribute to an overall engaging experience. Next, we elaborate on our design approach by considering three important aspects of character: backstory, social verbal and nonverbal behaviors, and expression of culture.

Backstory

Backstory is the history of a character, used to provide depth to the character (as it is incrementally revealed to users) and as a context for the character's actions. To create characters with consistent backstories, we have been collaborating with faculty and students from Carnegie Mellon's School of Drama. To begin, the drama personnel brainstorm several candidate characters and we jointly choose one that we believe would be both interesting and acceptable to our "audience": visitors and people in our buildings.

Once a character is chosen, the drama personnel develop a backstory and evolving story line (along with a three-dimensional graphical head that exemplifies that character, which we animate using our adaptation of SimpleFace). The backstory includes details about the character, including its age, family, living situation, past employment, and so on. The drama personnel write robot dialogue for each aspect of the backstory and the dialogues are entered into a content database.

The story line consists of 3–4 narrative threads that run throughout an academic year. The threads are scheduled out in advance, with different events occurring at different dates, each of the threads having a dramatic structure — building to a climax and resolution. Most of the story lines last the full year, but some end after only a few months, with others starting in their place. Examples of story lines that we have used include family problems, problems at the job (for example, thinking that the robot's supervisor is out to get him/her), pursuing a musical career, trying to lobby for robot rights, helping an Afghan refugee, and the difficulties of being a robot without arms. The drama personnel write dialogue for the character, describing both its pre- and postreaction to the events of the story lines. For instance, if the robot has a date scheduled for Saturday, the script might indicate that the robot can start talking about the date three days prior, at which point it might reveal that it is very excited and nervous. Then, after the date, different dialogue may be indicated, for instance if the date did not go well. Typically, a full year's worth of dialogue is written at one time and entered into the content database. In addition, specific nonverbal expressions can be designed and associated with pieces of dialogue to enhance their effect.

The content database associates dialogues (and expressions) with objects involved in the backstory and story lines, such as characters, buildings, pets, and nearby knickknacks. Each object has a list of attributes (for example, name, age, parent, job) and a list of associated events that involve that object. Each attribute is a *time line*, partitioned into a set of intervals over which a particular piece of dialogue holds. For instance, the "boyfriend" attribute of Valerie might have one value from November 2003 to December 11, 2003, then another value until March 15, 2004, then another value from March 15, 2004 until the "end of time." When someone asks Valerie "who is your boyfriend," the system looks up the "Valerie" object from the database, accesses the "boyfriend" attribute, searches the time line to find the interval that includes the current date and time, and then says the associated dialogue and displays any associated expressions. Similarly, one can ask temporal questions (for example, "who was your boyfriend last January") and questions with embedded references (for example, "what is your pet's name," which involves looking up the current "pet" attribute of the robot and then looking up the current "name" of that object). For consistency, static
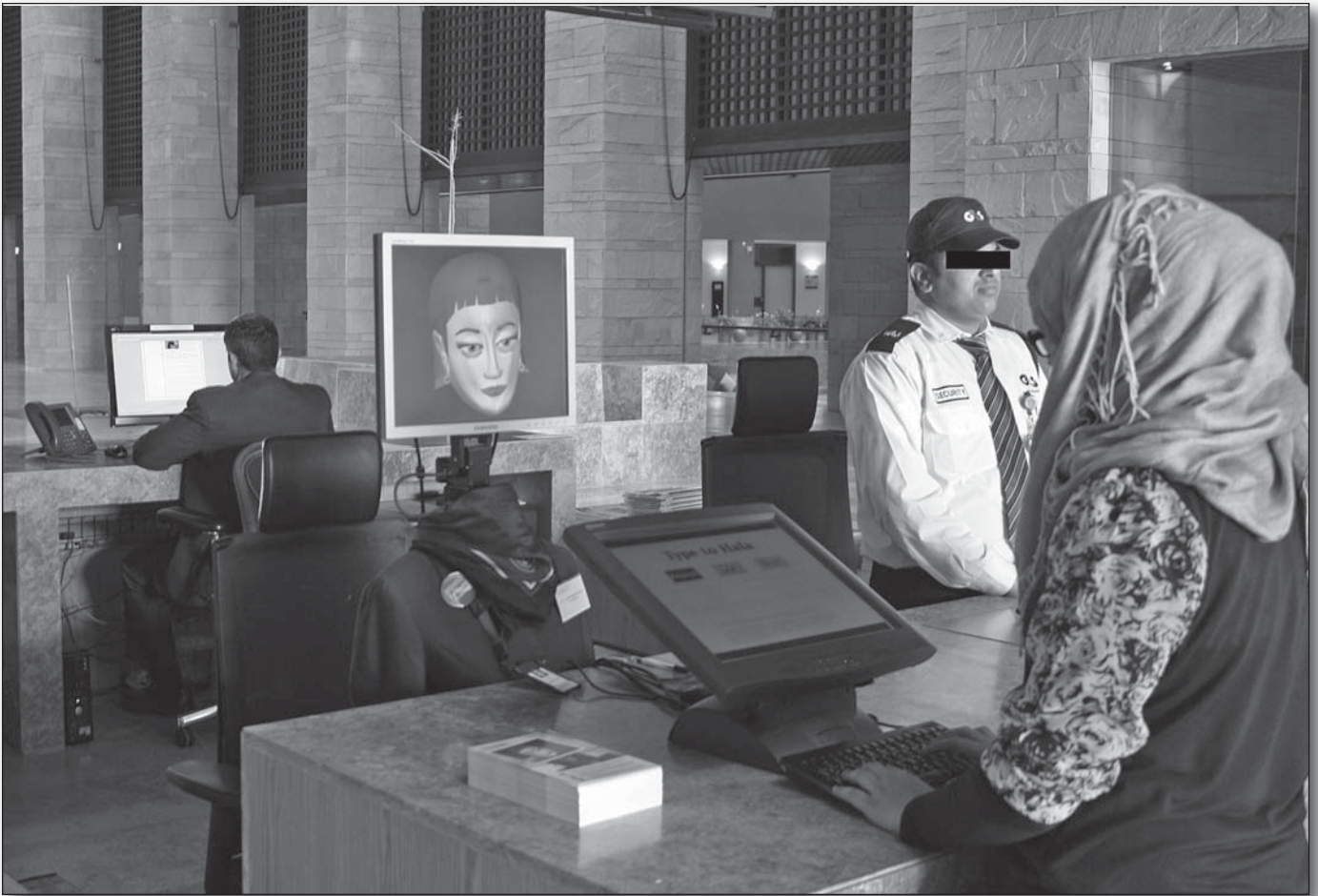
Figure 3. Hala in an Atrium with Human Receptionist and Security Staff.

information, such as the robot's name, is also entered into the database as an attribute with just a single interval on the time line. A scripting language was developed to facilitate expressing the temporal aspects of the story lines and backstory, which are parsed and entered into the database each time the robot software is restarted.

Our experience has shown that much care is needed to present the desired characters, and that the character development must be holistic. One pertinent example is with Valerie — a mid-20s female character with mother and boyfriend issues and a strong desire to become a lounge singer. Valerie was conceived of as being naïve and slightly neurotic, but very pleasant. Unfortunately, one of her responses to not being able to parse user input was seen as snide, or even hostile: "Look, I can give you directions, or sing you a song, or give you the weather. But, if you want anything else, I can't help you." Since that phrase was said often (Valerie fails to understand about a quarter of user input) it tended to dominate people's perception of her, in a way contrary to Valerie's intended personality. For Tank, we made a concerted effort to

have all such phrases reflect his character and personality.

Social Verbal and Nonverbal Behaviors

Exhibiting appropriate social behaviors is a large factor in the believability of characters. For robots in the wild, attracting interactors is an important goal, and we believe that socially appropriate behaviors can facilitate that. For instance, the roboceptionists greet passersby, either nodding at them or saying "good morning (afternoon or evening)," depending on the time of day. To avoid bothering people who are probably not interested in paying attention to the robot, we trained a classifier, using several weeks of collected laser range data, to predict how likely a person, at a given position and velocity, is to stop and interact with the robot. This classifier is run repeatedly as people are tracked by the laser, and when the confidence that a person will stop crosses a threshold, the robot greets that person.

Similarly, if a robot is idle and detects a person stopped nearby, but not in front of the keyboard, it will encourage the person to interact. If the per-
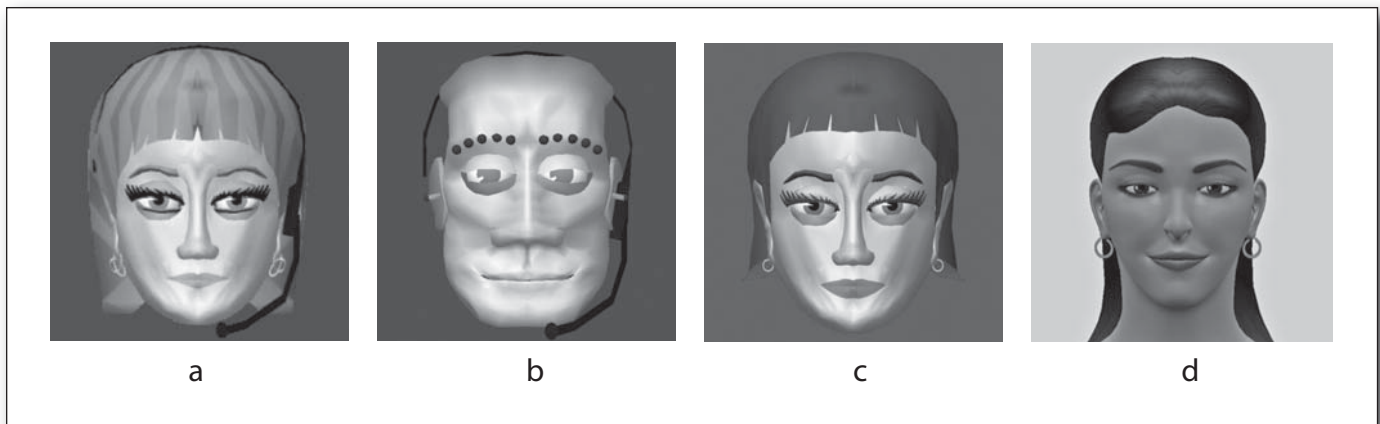
Figure 4. Faces of Roboceptionist Characters Valerie (a), Tank (b), Current Hala (c), and New Hala (d).

son is at the keyboard, but not typing, it will briefly describe how to interact with the robot. If the person continues to ignore the robot's requests to interact, the robot will announce that it is ignoring the person, too.

During the interaction, several strategies are used to encourage the user's social response. For example, the robot may augment the usual greetings with the phrase "Thanks for stopping by." We hypothesize that this primes the user to thank the robot more often. The robot also can indicate the effort it takes to answer the user's question, either by just pausing or saying, "Just a second, please. I'll look that up." The effects of these social strategies on the interactions are described in the evaluation section.

Nonverbal behaviors are just as important as dialogue to the design of believable characters. Facial expressions, gaze, and head posture all contribute to the overall effect of "liveness" and awareness of surroundings. We have developed a GUI that enables the drama personnel (and others) to design specific nonverbal expressions and associate them with phrases, so that when the dialogue is said the expressions are displayed.

In addition, we have implemented a categorical model of emotions (Gockley, Forlizzi, and Simmons 2006; Kirby, Forlizzi, and Simmons 2010) that enables the robot's facial expressions to be correlated both with the current interaction and the story lines. Developers can tag pieces of dialogue to indicate its emotional content. The emotional indicators are combined and used to change the expression of the robot. Specifically, we implemented a subset of the basic emotions presented in Ekman (1969): joy (happiness), sadness, disgust (frustration), and anger. The emotional expressions are based on Delsarte's code of facial expressions (Stebbins 1886), and their intensities are expressed over a continuous range with linear interpolation of key frames (figure 5). A web-based study demon-

strated that people were able to "read" the emotional expressions fairly readily (Kirby, Forlizzi, and Simmons 2010). Emotions are associated primarily with interactions and are short-lived, lasting the duration of the associated pieces of dialogue. For instance, the robot displays frustration if it does not understand an input, and when the input is "I love you," it displays happiness while responding, "But, you don't even know me."

The robot also maintains a longer-lived mood, which is primarily associated with personal history and "life" events. Mood is represented with a valence (positive or negative) and an intensity. The mood associated with an event rises and falls exponentially over time, reaching a peak at the time the event is scheduled to occur in the story line. The contributions of multiple concurrent events sum to produce the overall mood. In our model, mood affects emotions, and vice versa. Emotions are linearly scaled by the intensity of the current mood, where emotions that have the same valence as the mood are increased in intensity, while emotions that differ in valence are decreased. Similarly, the occurrence of emotional events can modulate mood. In accordance with existing psychological models (Rook 2001), positive social exchanges increase positive moods, while negative exchanges decrease any mood. Thus, expressing admiration for the robot will increase its already positive mood, while swearing at the robot can "bring it down." This effect decays over time, however, returning the robot to its "baseline" mood, as indicated by current events in the story line, if no further emotional events occur. The evaluation section describes an experiment to test the effects of mood on people's interactions with the roboceptionist.

The robots perform several other nonverbal behaviors to increase their believability. Periodically (and stochastically), they engage in autonomic behaviors — blinking, breathing (displayed through flaring nostrils), and small head move-
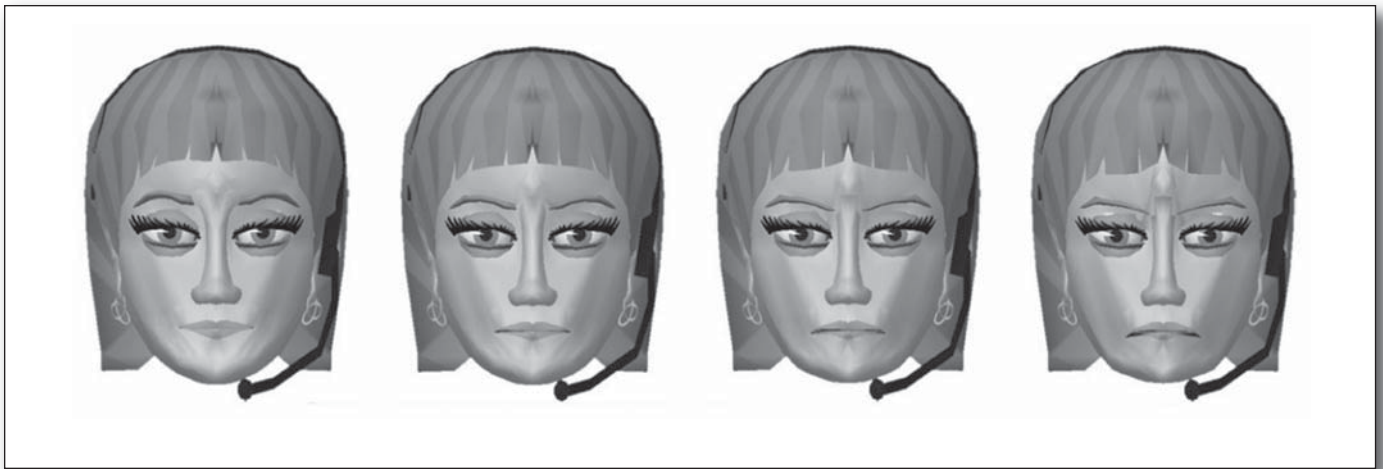
Figure 5. Valerie's Emotional Range from Neutral to Angry.

ments. Each adds a measure of liveness. The robots also use gaze to maintain engagement — while they focus primarily on the person at the keyboard, they will turn and nod at people who newly arrive and will occasionally thereafter turn to gaze at them. In addition, when someone is typing, the robots will periodically gaze down at the keyboard, to acknowledge that they are aware of the input.

Expressing Culture

Advocates of the holistic design of animated agents argue that cultural, as well as individual, variability should be expressed in all of the "qualities that animate characters should possess" (Hayes-Roth, Maldonado, and Moraes 2002), including backstory, appearance, manner of speaking, and gesturing. In particular, projecting a cultural identity may increase believability through a phenomenon known as *homophily* — the tendency of individuals to associate disproportionally with similar others (Lazarsfeld and Merton 1954). While homophily is typically observed in interactions between humans, its effects are also evident in interactions between humans and onscreen agents (Nass, Isbister, and Lee 2000). Based on this, we believe that affinity toward robot characters can be increased through an expression of cultural cues, especially when they are congruent with the culture of the user.

Most existing robot characters with cultural identity, such as Ibn Sina (Mavridis and Hanson 2009) and Geminoids (Ishiguro 2005), express their ethnicity almost exclusively through appearance and choice of language. While we attempted to avoid ethnic cues when designing the appearances of the Pittsburgh-based roboceptionists,[3] the new Hala (figure 4d) is decidedly Middle-Eastern in appearance and both versions of Hala support dialogue in Arabic and English.

To enhance cultural identity, we go beyond expressing ethnicity through appearance and language choice alone. Tank, for example, is aware of local colloquialisms (for example, "yinz" in Pittsburgh dialect is plural for "you") and Hala can handle some degree of code switching (for example, using the Arabic "inshallah" — "God willing" — in English dialogues). The roboceptionists' backstories support their identity as citizens of either Pittsburgh or Doha. Tank, for example, is an avid fan of the Pittsburgh Steelers ("I love the Steelers. A football team named after the most important material in the world is OK with me.") and once dated the Pittsburgh stadium's scoreboard; Hala is wary about the idea of driving in Doha's fast traffic.

In general, though, creating a believable character that projects a particular culture is quite challenging. The first difficulty lies with the term *culture* itself. When used to describe communities of people, it is as impossible to separate culture from language (Agar 1994) as it is to outline culture based solely on ethnicity or mother tongue (McPherson, Smith-Lovin, and Cook 2000). The more careful view of an individual's culture as a combination of dimensions that evolve with time and context and are viewed in a relation to an observer (Agar 2006) is alluring but implies a methodological difficulty in identifying culturally specific behaviors. While reported anthropological studies cover many world cultures to various extents, it is not always clear if the findings stand the test of time. It is also not unusual for an outsider's view of a particular community to be found offensive by the community members themselves (de Rosis, Pelachaud, and Poggi 2004).

To address these concerns more fully, we are developing a data-driven approach to identifying behaviors that express culture. First, we shortlist potential culturally specific behaviors, or *rich points* (Agar 1994), from anthropological accounts,

Topic	Fraction of Interactions (percent)
Location of a person/place	30.5
Weather	11.2
Date and time	2.5
Talking about the robot	31.0
Talking about the user	3.6
Greeting/farewell only	20.8
Gibberish	9.1
Insults	4.0
Others	2.5

Table 1. Frequency of Topics.

ethnographies, and studies on second-language acquisition, among others. We then design stimuli that incorporate these rich points and evaluate perception of those stimuli by the members of the desired demographics. For instance, in the case of an Arabic character, we define the community of interest based on native language and country of residence, and we compare the perception of the stimuli of that community with a control group (native English speakers residing in the United States).

Our first such study, which recruited participants from both American English and Arabic language communities using Amazon's Mechanical Turk, evaluated perceived personality and naturalness of linguistic features of verbosity, hedging, syntactic and lexical alignment, and formality through fragments of dialogues (Makatchev and Simmons 2011a). The results have been quite informative. For example, we found that the speakers of American English, unlike the speakers of Arabic, find formal utterances unnatural in dialogue acts involving greetings, question-answer, and disagreement. Formal utterances also tended to be perceived as indicators of openness and conscientiousness by Arabic speakers in disagreements and apologies, respectively, but not by American English speakers. Apologies that included the hedging marker "I am not sure" were perceived as an indicator of agreeableness by American English speakers, but not by speakers of Arabic. Additional studies in both verbal and nonverbal behaviors, as well as transfer of the findings to our robot test beds, are in progress.

Evaluating Robot Characters in the Wild

One of the consequences of deploying a robot in the wild, as opposed to studying the interactions through controlled laboratory experiments, is that it becomes difficult to survey and interview users without disrupting some of the "wilderness" — the impression that the interaction is autonomous and unmonitored by humans. As a result, corpora of interactions with robots that operate in the wild typically have neither user demographic data nor the ground truth of user intentions and impressions. On a positive side, robots operating in the wild may encounter many more users than in a reasonably controlled experiment. Tank, for example, even after the novelty factor wore out, still averages 27.5 interactions per day ($SE = 0.5$), each averaging 9.5 total dialogue turns ($SE = 0.1$) (based on January 2009 to June 2011 data). The relatively large amount of collected data makes it feasible to analyze multiple features of interactions at once. Specifically, we demonstrate that mining interaction transcripts allows us to infer the degree of the user's social orientation toward the robot. The data-driven method also enables us to estimate the joint effects that multiple robot behaviors have on the interactions. We also show that embedding elicitation questions in the interaction itself is a viable way of obtaining ground truth for some of aspects of the users' intentions.

Topics, Greetings, and Discourse Features

One assertion in the previous section is that the robots' backstories lead to more believable characters. Anecdotally, we know that many users frequent the robot to follow its stories, and some become quite emotionally attached. More quantitatively, we can analyze the topics of user interactions to indicate how people interact with the robot. Table 1, adopted from Lee and Makatchev (2009), shows the frequency of dialogue topics in interactions with Tank. These data were obtained by manually coding 197 interactions that occurred in 5 weekdays in March 2008 (2 coders, $\kappa = 0.7$). Interaction boundaries are defined by the user's approach and departure, according to the laser tracker.

While a large fraction of the interactions are seeking information on directions, weather, or time (42.7 percent), a comparable number are concerned with the robot's character itself (31.0 percent). Only about 12.7 percent of dialogues have more than one topic associated with them, and about 20.8 percent of all dialogues consist only of a greeting and an occasional farewell.

Little overlap (6.6 percent) between dialogues with information-seeking questions and those involving the robot's character suggests that there are (at least) two classes of users: those who are mainly interested in finding out information pertinent to their daily agendas and those who are curious about the robot.
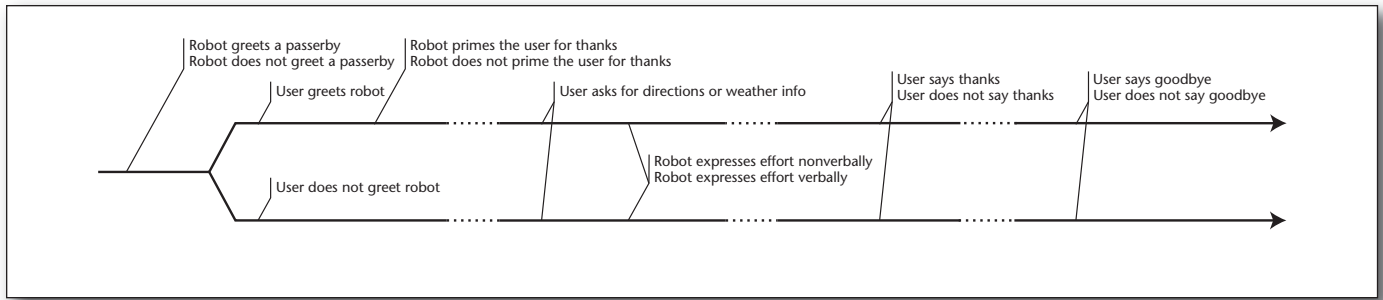
Analysis of dialogue openings provides more

Figure 6. The Temporal Flow of an Information-Seeking Interaction.

support for this hypothesis. The presence of a greeting in the user's first turn appears to be a strong predictor of many features of the following dialogue (Makatchev, Lee, and Simmons 2009). Users who start with a greeting tend to be more persistent in response to the robot's nonunderstandings (77 percent versus 67 percent of users who did not greet), more than three times as likely to thank the robot for an answered question (25 percent versus 8 percent), and more likely to end their interactions with a farewell (20 percent versus 13 percent). The users who greeted the robot also perform better on the information-seeking task: 54 percent of them have all of their utterances successfully parsed (versus 44 percent) and 50 percent of them get their questions answered (versus 43 percent). On the other hand, the verbosity of the dialogue, excluding the greeting and farewell turns, is relatively unchanged, compared to users that did not greet.

Similar associations between dialogue openings and discourse features, such as lexical and syntactic alignment, clarification questions, reformulations, and repetitions, have been found in various HCI and HRI corpora by Fischer (see, for example, Fischer [2006]). Fischer explains these correlations by the user's preconceptions about computer agents and robots. In particular, those users who consider the robot as a tool will neither align nor produce various social behaviors, while those who consider the robot as a conversational partner will. Similarly, Lee, Kiesler, and Forlizzi (2010), based on an analysis of the roboceptionist data, suggest that users who have different mental models of the robots will apply different interaction scripts.

A similar analysis was performed on Hala's corpus, to see what differences may be observed between a Pittsburgh-based and Doha-based robot. Due to differences with respect to their knowledge base and dialogue rule coverage, a direct comparison between their corpora is difficult. Nevertheless, we conducted such a comparison (Makatchev et al. 2010), attempting to normalize Hala's corpus by focusing on those dialogues that were conducted entirely in English.

The comparison shows that Hala's dialogues last almost twice as long as Tank's (120 seconds versus 63 seconds) and on average contain one extra pair of utterances. The fraction of dialogues that start with a greeting are about the same for Hala and Tank (38.7 percent and 39.4 percent). Tank's answers receive thanks at a much higher rate than Hala's (12.9 percent versus 2.3 percent), perhaps in part because more of Hala's questions are of a personal nature (57 percent versus Tank's 31 percent), such as "Are you married?" Other possible reasons for the observed differences include distinct robot personae, perceived gender differences, differences in coverage of their knowledge bases, their immediate surroundings, and potential differences in demographics of their user communities (Fanaswala, Browning, and Sakr 2011).

Eliciting Social Behaviors through Dialogue Strategies

The results suggest that users who produce social dialogue acts are also more likely to get their questions answered by the robot. Thus, we would like to see whether we can develop robot interaction strategies that encourage users to behave more socially toward the robots, and whether by doing so we can influence task performance.

Toward this end, we have investigated the use of *initiative*, *priming,* and *expression of effort* to elicit users' social responses. Figure 6 depicts the dialogue structure in terms of these strategies and possible user responses.

Initiative: An interaction with the roboceptionist may begin in one of two ways. It can be initiated by the robot, when the robot's laser tracker detects a passerby, and consists of either a greeting, a head turn and nod, or a combination. Alternately, an interaction can be initiated by the user, by typing to the robot. Our analysis (Makatchev, Lee, and Simmons 2009) shows that users who engage in a dialog with the robot after being greeted first (with a verbal, nonverbal, or combined greeting) start their interactions with a greeting more often than users who were not proactively greeted (42 percent versus 37 percent). In addition, in support of our

hypothesis, users who were greeted by the robot were also more likely to have their questions successfully parsed and answered (52 percent versus 42 percent).

Priming: In an attempt to prime the user to say "thanks" after the robot's answers, we evaluated having the robot preface its standard response to a user's greeting with the phrase "Thanks for stopping by." We found that such priming had a significant positive main effect on occurrences of social dialogue acts of thanking and farewells (Makatchev and Simmons 2011b). In particular, logit model selection with second-order interaction terms results in a model with the priming variable significant with $p = 0.03$ (odds ratio's 95 percent confidence interval is [1.06, 3.15]) for explaining occurrences of thanks and with $p < 0.01$ (odds ratio's 95 percent confidence interval is [1.16, 1.64]) for explaining occurrences of farewells.

Expression of Effort: We hypothesize that if the robot indicates that it has put effort into answering a user's question, the user may be more likely to provide a social response. To test this, we had Tank randomly precede its answers to weather and direction questions with either a silent half-second pause (a nonverbal expression of effort, or NVEE) or one of the following two phrases (verbal expressions of effort, or VEE): "Wait a second please. I'll look that up" or "I am looking it up. Please hold on." We found a small positive main effect of NVEE on the turnwise length of the interactions. There are, however, some significant interaction effects. For example, logistic regression on counts of occurrences of expression of effort indicates that users who greeted the robot were more likely to express thanks and farewells as the number of occurrences of NVEE increased ($p < 0.01$, odds coefficient's 95 percent confidence interval is [1.19, 3.08]). The VEEs did not produce this kind of effect. One possible explanation is that multiple occurrences of a VEE may be annoying. Direct comparison of NVEE and VEE shows two interesting, but weakly significant ($0.05 < p < 0.1$), interactions. First, VEE in combination with the robot's admission of inability to parse the user's utterance has more chance of being thanked by the user than NVEE co-occurring with such failures. Conversely, NVEE combined with a valid robot answer has more chance of being followed by a user's thanks and farewell than VEE combined with a valid answer.

In summary, initiative in greeting users correlates with having them exhibit social attitudes toward the robot. Priming with "Thanks for stopping by" succeeds in encouraging users toward social dialogue. Multiple nonverbal expressions of effort improve the rates of social dialogue acts for users who have already greeted the robot. All users,

on average, produced more social dialogue acts in response to nonverbal expressions of effort combined with a valid answer or verbal expressions of effort combined with a failure to parse or fetch information.

Expressing Emotion and Mood

In a similar fashion, we would like to see how nonverbal expressions of emotion and mood affect users' social attitudes toward the robots. While previous research demonstrated that emotions displayed by robots can be recognized and interpreted by humans (Breazeal 2003), it was not known how people would react to a *mood* (that is, a consistent display of positive or negative emotion).

Gockley, Forlizzi, and Simmons (2006) tested the model of emotion and mood described previously. The study displayed the mood of Valerie (sad, happy, or neutral) as a combination of facial expressions and several nonverbal behaviors: In the negative (sad) condition, Valerie either looked away from the visitor or appeared to sigh. In the neutral condition, Valerie either smiled and performed a single head nod or briefly glanced away from the visitor. In the positive (happy) condition, Valerie either smiled and nodded (as in the neutral condition, but with a wider smile), or bounced her head from side to side in a seemingly happy, energetic motion.

During the nine-week-long study, each day the robot displayed either a negative, neutral, or positive mood. For consistency, the mood was coordinated with the robot's story line.

The analysis was performed separately for weeks with low and high visitor traffic. Dialogues during low-visitor traffic weeks had more user turns in the neutral condition than in either the positive or negative conditions (neutral $M = 4.19$, positive $M = 3.49$, negative $M = 3.74$; $F[1, 1236] = 4.59$, $p = 0.03$). Positive and negative mood conditions did not differ significantly.

During high-visitor traffic weeks, however, dialogues had more user turns in the positive and negative mood conditions than in the neutral condition (neutral $M = 3.33$, positive $M = 3.92$, negative $M = 3.79$; $F[1, 1419] = 3.81$, $p = 0.05$). Again, there were no significant differences between positive and negative mood conditions.

A possible explanation for these opposing trends is that high-visitor traffic weeks consist of more first-time visitors, who may find it more interesting to encounter a moody (sad or happy) robot. On the other hand, repeat users (more likely during low-traffic weeks) may perceive a higher degree of common ground in the moody (nonneutral) conditions, which may lead to more efficient interactions. More detailed analysis of the interactions would be necessary to give a conclusive answer on the effect of the nonverbal expression

of the mood on the social perception of the robot. However, the observed effect of the robot's mood on the number of dialogue turns suggests that factors related to social perception of the robot can be influenced by the robot designers.

Pragmatic Analysis

Humans often produce utterances that convey intentions beyond those implied by the syntax, or even semantics. For example, some apparent questions do not seek information, but instead are intended as either assertions (*rhetorical questions,* such as "Who do you think you are?"), social rituals (*phatic questions,* such as "How are you?"), or tests of the receiver's knowledge (*display questions,* such as "Where is my office?"). If a robot character does not seem to understand the deeper, *pragmatic meaning* (meaning that depends on context), its degree of believability may be lessened in the estimation of its users.

Consider the fragment of a dialogue shown in figure 7. Clearly, the user has some knowledge of where Jane's office is (the robot's answer is actually incorrect). Why did the user ask this question to the robot in the first place? What was the user's intention? Is this the best possible response, for the robot to provide an answer referring to the semantic content of the query? While we do not yet have answers to these important questions, here we address a problem essential for development of display question detection methods: obtaining ground-truth display question annotations.

The roboceptionist dialogues occasionally include fragments where users provide explicit feedback, or even answers to their own questions, after the robot fails to answer them properly. For example, dialogues such as the one in figure 7 suggest that users do ask display questions and that some users are willing to provide detailed feedback after the robot fails to produce a satisfactory answer. We attempt to exploit these tendencies by having the robot purposely fail to answer questions at random and, instead, present an *elicitation question* (but not more than once per dialogue). A fragment of such a dialogue is shown in figure 8. The subset of display questions that is labeled this way serves as a lower bound on the fraction of interactions that contain a display question among all the interactions with an elicitation question. This is just a lower bound because the user may choose not to answer the elicitation question even if he/she knows the answer, and the elicitation intervention may be applied to a question that is not a display question within a larger dialogue that does contain a display question.

An experiment conducted over a period of 3 months shows that at least 16.7 percent (*SE* = 4.6 percent) of interactions that included an elicitation question contained a display question

1 U: Jane Smith

2 R: Jane Smith is in 4001 Wean Hall.
    Would you like directions?

3 U: no she is not1

4 R: OK

5 U: Smith is not where you say she is

Figure 7. A Fragment of a Dialogue Containing User Feedback.

The person's name and room are modified for privacy, and the labels "U:" and "R:" are added to denote the user and robot turns, respectively.

(Makatchev and Simmons 2010). Given the relatively large number of display questions, we believe that it is important to be able to reliably detect and handle such questions, in order to sustain believability. Future work will analyze the corpus to find ways of distinguishing between display and other types of questions.

Conclusions

When users encounter a robot capable of natural language dialogue, their expectations are raised not only with respect to the natural language understanding and generation capabilities, but also with regard to the intentionality of both verbal and nonverbal behaviors, the robot's autonomy, and its awareness of the sociocultural context. Failing to meet user expectations along any of these dimensions may result in less natural or even disrupted interactions.

We attempt to meet such expectations by creating believable characters that provide an illusion of life. We presented several approaches to creating believable robot characters, including having a backstory and dynamic story line, using nonverbal expressions of emotions, and incorporating social cues and sociocultural context into the robot's behaviors. Our results show that such a holistic approach to robot design results in robots that can sustain the interest of a user community over many years of deployment. We argue for the desirability of evaluating design choices "in the wild" and have shown that, in some cases, it is feasible to obtain ground-truth user intent without breaking the illusion of the robot operating unsupervised and unmonitored.

1 U: where is hamburg hall

2 R: Sorry. I could not get the information. If you happen to know the answer to your question, can you tell it to me?

3 U: yes

4 R: Ok, what is the answer?

5 U: go out of the double doors behind you and walk to your left, the path will lead you to the back of hamburg hall.

Figure 8. A Dialogue Fragment Containing the Embedded Elicitation Question in Turns 2 and 4.

We also found that many users communicate with the robot characters using social dialogue, which can be considered as a measure of believability. Such users tend to perform better on the information-seeking task. The natural question arises: is there a causal link between social attitude and performance? We took a step toward answering this question by finding that dialogue strategies of priming and expression of effort tended to increase the social responses of users who had already greeted the robot. In addition, there is weaker evidence that verbal expression of effort can trigger social dialogue acts for all users (on average) when the robot fails to answer the question. It remains to be shown, however, whether such a manipulation of users' social responses affects their task performance.

Not all tasks would benefit from social interactions with a believable agent. For example, for certain stressful and demanding tasks, even human teams use simplified languages devoid of social cues. Nevertheless, for tasks where social interaction is key, such as when a receptionist establishes rapport with a visitor, we contend that believable robot characters have the potential to positively effect both perceptual and performance metrics of interactions.

Future work includes expansions along all of the outlined dimensions of believability, including verbal and nonverbal expression of personality and emotion, as well as recognition and generation of culturally appropriate behaviors. While we still have a long way to go in understanding exactly what factors affect character believability, we feel confident that our approach is heading in the right direction.

Acknowledgments

Note

1. See Cepstral LLC, Cepstral Text-to-Speech (cepstral.com).

2. See the Ainebot project home page: distro.ibiblio.org/pub/linux/distributions/amigolinux/download/ainebot.

3. However, the new Pittsburgh roboceptionist character (under development) will have an African-American appearance and a Southern accent.

References

Agar, M. 1994. *Language Shock: Understanding the Culture of Conversation*. New York: William Morrow.

Agar, M. 2006. Culture: Can You Take It Anywhere? Invited Lecture Presented at the Gevirtz Graduate School of Education, University of California at Santa Barbara. *International Journal of Qualitative Methods* 5(2): 1–12.

Bates, J. 1994. The Role of Emotion in Believable Agents. *Communications of the ACM* 37(7): 122–125.

Breazeal, C. 2003. Emotion and Sociable Humanoid Robots. *International Journal of Human Computer Studies* 59(1–2): 119–155.

de Rosis, F.; Pelachaud, C.; and Poggi, I. 2004. Transcultural Believability in Embodied Agents: A Matter of Consistent Adaptation. In *Agent Culture. Human-Agent Interaction in a Multicultural World,* ed. S. Payr and R. Trappl, 75–105. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Ekman, P. 1969. Pan-Cultural Elements in Facial Displays of Emotion. *Science* 164(3875): 86–88.

Fanaswala, I.; Browning, B.; and Sakr, M. 2011. Interac-

tional Disparities in English and Arabic Native Speakers with a Bi-Lingual Robot Receptionist. In *Proceedings of the 6th International Conference on Human-Robot Interaction (HRI),* 133–134. New York: Association for Computing Machinery.

Fischer, K. 2006. The Role of Users' Preconceptions in Talking to Computers and Robots. Paper presented at the Workshop on How People Talk to Computers, Robots, and Other Artificial Communication Partners, Delmenhorst, Germany, 21–23 April.

Gockley, R.; Bruce, A.; Forlizzi, J.; Michalowski, M.; Mundell, A.; Rosenthal, S.; Sellner, B.; Simmons, R.; Snipes, K.; Schultz, A. C.; and Wang, J. 2005. Designing Robots for Long-Term Social Interaction. In *Proceedings of the 2005 IEEE/RS International Conference on Intelligent Robots and Systems,* 2199–2204. Piscataway, NJ: The Institute of Electrical and Electronics Engineers.

Gockley, R.; Forlizzi, J.; and Simmons, R. 2006. Interactions with a Moody Robot. In *Proceedings of the 5th ACM/IEEE International Conference on Human Robot Interaction,* 186–193. New York: Association for Computing Machinery.

Hayes-Roth, B.; Maldonado, H.; and Moraes, M. 2002. Designing for Diversity: Multicultural Characters for a Multicultural World. Paper presented at IMAGINA, The European 3D Simulation and Virtual Technology Event. Monte Carlo, Monaco, 12–15 February.

Hoffman, G. 2011. On Stage: Robots as Performers. Paper presented at the RSS 2011 Workshop on Human-Robot Interaction: Perspectives and Contributions to Robotics from the Human Sciences. Los Angeles, CA, 1 July.

Ishiguro, H. 2005. Android Science. Toward a New Cross-Interdisciplinary Framework. Paper presented at Toward Social Mechanisms of Android Science: An ICCS/CogSci-2006 Long Symposium. Vancouver, BC, Canada, 26 July.

Kirby, R.; Forlizzi, J.; and Simmons, R. 2010. Affective Social Robots. *Robotics and Autonomous Systems* 58(3): 322–332.

Lazarsfeld, P., and Merton, R. 1954. Friendship as a Social Process: A Substantive and Methodological Analysis. In *Freedom and Control in Modern Society*, ed M. Berger, T. Abel, C. Page. New York: Van Nostrant.

Lee, M. K., and Makatchev, M. 2009. How Do People Talk with a Robot? An Analysis of Human-Robot Dialogues in the Real World. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, 3769–3774. New York: Association for Computing Machinery.

Lee, M. K.; Kiesler, S.; and Forlizzi, J. 2010. Receptionist or Information Kiosk: How Do People Talk with a Robot? In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*, CSCW, 31–40. New York: Association for Computing Machinery.

Makatchev, M., and Simmons, R. 2010. Do You Really Want to Know? Display Questions in Human-Robot Dialogues. In *Dialog with Robots: Papers from the AAAI Fall Symposium.* Technical Report FS-10-05. Menlo Park, CA: AAAI Press..

Makatchev, M., and Simmons, R. 2011a. Perception of Personality and Naturalness through Dialogues by Native Speakers of American English and Arabic. Paper presented at the 12th Annual SIGdial Meeting on Discourse and Dialogue. Portland, OR, 17–18 June.

Makatchev, M., and Simmons, R. 2011b. Using Initiative,

ICWSM-12

Submissions Due January 18, 2012

The Sixth International AAAI Conference on Weblogs and Social Media (ICWSM-12) will be held at Trinity College in Dublin, Ireland, June 4-8, 2012. This unique forum brings together researchers from the disciplines of computer science, linguistics, communication, and the social sciences. The broad goal of ICWSM is to increase understanding of social media in all its incarnations. Submissions describing research that blends social science and computational approaches are especially encouraged.

In addition to the usual program of contributed technical talks, posters and invited presentations, the main conference will include a selection of keynote talks from prominent social scientists and technologists. Keynotes will be presented by Andrew Tomkins (Google+), Patrick Meier (Ushahidi!), and Lada Adamic (University of Michigan). The very successful workshop and tutorial programs will continue on the first day of the conference, Monday, June 4.

Please note the following important deadlines for ICWSM-12:

January 6	Workshop Proposal Acceptance
January 9	Tutorial Proposal Submission
January 13	Paper, Poster, Demo Abstract Submission
January 18	Full Paper, Poster, Demo Submission
January 23	Tutorial Proposal Acceptance
February 27	Paper, Poster, Demo Notifications
March 12	Camera Ready Paper Due
June 4-8	ICWSM-12 Conference

For complete submission instructions, deadlines, and other details about the conference, please see www.icwsm.org/2012 or write to icwsm12@aaai.org

Priming, and Expression of Effort to Elicit Social Behaviors in a Dialogue with a Robot Receptionist. Technical Report TR-11-23, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.

Makatchev, M.; Fanaswala, I. A.; Abdulsalam, A. A.; Browning, B.; Gazzawi, W. M.; Sakr, M.; and Simmons, R. 2010. Dialogue Patterns of an Arabic Robot Receptionist. In *Proceedings of the 5th ACM IEEE International Conference on Human-Robot Interaction*, 167–168. New York: Association for Computing Machinery.

Makatchev, M.; Lee, M. K.; and Simmons, R. 2009. Relating Initial Turns of Human-Robot Dialogues to Discourse. In *Proceedings of the 4th ACM IEEE International Conference on Human-Robot Interaction*, 321–322. New York: Association for Computing Machinery.

Mavridis, N., and Hanson, D. 2009. The IbnSina Interactive Theater: Where Humans, Robots and Virtual Characters Meet. In *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication.* Piscataway, NJ: The Institute of Electrical and Electronics Engineers.

McPherson, M.; Smith-Lovin, L.; and Cook, J. M. 2000. What Is a Language Community? *American Journal of Political Scienc*e 44(1): 142–155.

Nass, C.; Isbister, K.; and Lee, E. 2000. Truth Is Beauty: Researching Embodied Conversational Agents. In *Embodied Conversational Agents,* ed. J. Cassell, J. Sullivan, S. Prevost, and E. Churchill, 374–402. Cambridge, MA: MIT Press.

Rook, K. S. 2001. Emotional Health and Positive Versus Negative Social Exchanges: A Daily Diary Analysis. *Applied Developmental Sciences* 5(2): 86–97.

Stanislavski, K. 2008. *An Actor's Work: A Student's Diary.* New York, NY: Routledge.

Stebbins, G. 1886. *Delsarte System of Dramatic Expression.* New York: E. S. Werner.

Reid Simmons is a research professor in the Robotics Institute at Carnegie Mellon University. He received his Ph.D. in artificial intelligence from MIT in 1987. His research focuses on developing reliable autonomous systems that can plan, act, and learn in complex, uncertain environments. He is currently investigating aspects of human-robot social interaction, both conversational and navigational, multirobot coordination, especially as it applies to assembly of large-scale structures, and assistive robotics to aid the elderly and people with disabilities.

Maxim Makatchev is a Ph.D. candidate in the Robotics Institute at Carnegie Mellon University. His research interests are in human-robot interaction and natural language dialogue.

Rachel Kirby (nee **Gockley**) received her Ph.D. in robotics from Carnegie Mellon University in 2010. Her research addressed various aspects of social human-robot interaction, with a focus on computational models of human behavior. Her thesis "Social Robot Navigation" presented a method for robots to navigate safely and comfortably around people by respecting human social conventions, such as personal space. Kirby is now a software engineer at Google, Inc.

Min Kyung Lee is a Ph.D. student in the Human-Computer Interaction Institute at Carnegie Mellon University. Her general research interests lie in the broad areas of human-computer interaction, human-robot interaction, and computer-supportive collaborative work. Over the past four years, Kyung has worked on several projects that focused on designing social behaviors of autonomous systems and their application in the domain of collaboration, and personal and assistive services at Intel and Willow Garage.

Imran Fanaswala is a senior research programmer at Carnegie Mellon Qatar. His time is spent coding, mentoring, and mulling, in the areas of human-robot interaction and cloud computing.

Brett Browning is a senior system scientist in the Robotics Institute at Carnegie Mellon University. He also is associated with Carnegie Mellon Qatar, where he is a codirector of the Qri8 robotics lab, and the National Robotics Engineering Center (NREC). He joined Carnegie Mellon in 2000 as a postdoctoral fellow and then the faculty in 2002. Browning received his Ph.D. in electrical engineering from the University of Queensland in 2000. Browning's research interests focus on perception and autonomy for robots operating in industrial and off-road environments.

Jodi Forlizzi is an associate professor of design and human-computer interaction at Carnegie Mellon University, in Pittsburgh, Pennsylvania. Forlizzi is an interaction designer who examines theories of experience, emotion, and social product use as they relate to interaction design. Other research and practice centers on notification systems ranging from peripheral displays to embodied robots, with a special focus on the social behavior evoked by these systems. One recent system is Snackbot, a robot that delivers snacks and encourages people to make healthy choices.

Majd F. Sakr, Ph.D., is an associate teaching professor in computer science and the assistant dean for research at Carnegie Mellon University in Qatar. He is the cofounder of the Qatar Cloud Computing Center. In addition to working at Carnegie Mellon in Pittsburgh, he has held appointments at the American University of Science and Technology in Beirut and at the NEC Research Institute in Princeton, New Jersey. His current areas of research interest include data-intensive scalable computing and cross-cultural, multilingual human-robot interaction and HCI. He holds a BS, MS and Ph.D. in electrical engineering from the University of Pittsburgh.