Mixed-Initiative Long-Term Interactions with an All-Day-Companion Robot

Stephanie Rosenthal and **Manuela Veloso**
School of Computer Science
Carnegie Mellon University
Pittsburgh PA USA

Abstract

As robots become incorporated into our environments, they must be equipped with the ability to communicate effectively with us. In particular, robots that perform longer tasks for a small set of people (*e.g.,* a companion robot to escort visitors to meetings all day) need to be able to start and maintain interesting and relevant dialog with any and all humans involved. In this work, we present our ongoing work on our robot, CoBot, which is assigned an all-day task to escort a visitor around our building and perform tasks for her. We first describe CoBot's dialog manager which is responsible for the task-oriented dialog, including dialog to meet the visitor's needs, CoBot's notifications of interesting locations around the building, and the robot's own requests for help. We, then, focus two aspects of the dialog manager: 1) how CoBot can invoke more accurate answers to its requests for help from the visitor and 2) how to reduce repetitive dialog which can happen during all-day interactions. We provide an example dialog between CoBot and a visitor to illustrate the dialog manager's capabilities.

Introduction

As robots become incorporated into our environments to perform tasks (*e.g.,* in our offices (Asoh et al. 1997; Bohus and Horvitz 2009; Lee et al. 2009), museums (Thrun and others 2000), nursing homes (Pineau et al. 2003), or malls (Kanda et al. 2009)), they must be equipped with the ability to initiate and maintain task-oriented dialog to receive tasks from, coordinate actions with, and complete tasks for humans. In order to help the robots deal with the uncertainty in the environment and dialog as well as maintaining human interest in robots performing longer-term tasks for humans, robots should have support for 1) *mixed initiative* dialog to both receive and respond to task requests and 2) the ability to avoid repetitive dialog during long interactions.

To illustrate the all-day mixed-initiative relationship, we contribute the dialog manager of our robot, CoBot (Figure 1), that performs the Visitor-Companion Task. The Task requires a robot to escort a visitor to meetings throughout the day and provide other services such as getting coffee and

(a) front

(b) back

Figure 1: The CoBot Visitor-Companion Robot (Biswas and Veloso 2010; Rosenthal, Biswas, and Veloso 2010) (Hardware designed and built by Mike Licitra)

printing papers (see (Rosenthal, Biswas, and Veloso 2010) for complete details). Because the Visitor-Companion Task requires that a visitor be present near the robot for a majority of the time, CoBot can maintain some state of the visitor to proactively offer help in addition to handling the visitor's own requests for amenities. For example, because CoBot maintains information about when a visitor has received coffee, it can know both when the visitor might need coffee again as well as when it is too soon to offer.

Additionally, the visitor's proximity near the robot through the day provides an opportunity for the robot to ask for help to overcome its own limitations. CoBot, in particular may become uncertain of its location while navigating and it lacks arms so it cannot lift a coffee pot to pour coffee. While many other robots have requested help from humans, most questions are aimed at clarifying direct commands such as navigating an area (*e.g.,* (Fong, Thorpe, and Baur 2003; Breummer et al. 2005; Heger et al. 2005)) or looking up information (*e.g.,* (Gockley and others 2005; Bohus and Horvitz 2009)). Instead, CoBot requests help from humans who are not controlling the robot and may not know anything about how CoBot works. Because humans are better than the robot at determining the current location or physically manipulating objects like coffee mugs, the visitor can help the robot overcome its limitations which, in turn, helps the robot complete more tasks for the human.

In this work, we first describe our on-going work on CoBot's dialog manager which is responsible for the mixed-initiative task-oriented dialog throughout the day. This includes the visitor's task requests and Cobot's help offerings, CoBot's notifications about interesting labs and locations around the building, and CoBot's own requests for help. Then, we focus on how CoBot can invoke more accurate answers from the visitor while asking for localization help by changing the contextual information it provides. We show that visitors are most accurate when the robot not only provides verbal description of its location, but also an exact location prediction on a map, its uncertainty in that prediction, and asks the visitor to describe the location. Next, we describe our current and future efforts to reduce repetitive dialog that can happen during all-day dialogs. Finally, we provide an example dialog between CoBot and a visitor to illustrate the dialog manager's capabilities.

Mixed-Initiative Dialog

CoBot has the capability to interact with humans in the environment through a type-to interface on its laptop and through speech with a wireless headset microphone and speakers (Figure 1(b)). Because it can be difficult to type to a moving robot, it is essential that the robot understand basic speech. However, we understand that speech recognition is still difficult with accented voices and in noisy environments, so we provide the type-to interface to eliminate some of the potential frustrations with speech recognition.

CoBot's mixed-initiative dialog manager allows the visitor to make task-related requests (*i.e.,* request a drink or information about meeting hosts) and allows the robot to proactively ask if the visitor needs anything (Rich and Sidner 1998; Horvitz 1999). CoBot can also notify a visitor of important places of interest without requiring a response. Finally, CoBot can request help from the visitor or other humans in the environment if it becomes uncertain of its current state or if it cannot perform an action that the visitor requests. We enumerate and explain the CoBot's dialog manager as it pertains to the Visitor-Companion Task.

Visitor's Task Requests and Robot Task Offerings

The Visitor-Companion task requires that the robot provide information about the day's schedule, directions to restrooms and other locations, and actions on behalf of visitors or meeting hosts. CoBot knows the visitor's schedule and can share information about when meetings start, who the meetings are with, and where they are located. CoBot knows other important places in the building such as elevators, restrooms, and kitchens and can direct the visitor to them. Additionally, we have ongoing work towards the robot performing actions for the visitor such as getting coffee, printing research papers, and calling for a taxi.

Both the visitor and CoBot can initiate dialog on these topics. Because it is hard for the robot to predict most of the her needs, we rely on the visitor to make requests when she needs to. However, CoBot does maintain state about the visitor through the dialog manager (Rosenthal, Biswas, and Veloso 2010). Using the dialog history, CoBot can determine when it is appropriate to offer an amenity such as coffee if the visitor has not requested it recently. At the end of a scheduled meeting, the robot can also notify the visitor and ask if she is ready to navigate to the next meeting.

If the visitor makes a request that CoBot cannot answer, CoBot has the ability to perform internet searches and to request help from other humans through dialog or email. For example, if the visitor requests more information on their meeting hosts, CoBot has the ability to display the host's website for the visitor to read (we do not currently perform text-to-speech reading of websites). If the visitor is running late, CoBot can email the next meeting host to notify him. These interactions with the internet and other humans in the environment allow CoBot to be a larger source of information for the visitor and meeting hosts.

Robot's Notifications

CoBot has the ability to notify the visitor as she passes important or interesting locations and as it completes actions for her. For example, CoBot might tell a visitor with interests in robotics about the robotics labs that they pass in the halls. It might also share information about its own hardware and software. CoBot can also notify the visitor when it receives information about when and where to meet a taxicab if she requests one. These notifications do not require a response from the visitor and are meant only as additional information that keeps the visitor engaged in the interaction. Later, we will discuss our current and future efforts to reduce repetition of notifications throughout the visitor's day.

Robot's Help Requests

While our ultimate goal is for robots like CoBot to perform tasks autonomously, we realize that today's robots still have many limitations at the perception, cognition, and execution levels. For example, CoBot lacks arms so it could never pour coffee from a coffee pot autonomously and sometimes CoBot becomes uncertain of its location so it may require help to relocalize. Interestingly, many of the robots limitations are strengths for humans who coexist with robots in the environment. Humans can lift and pour coffee easily and read the office numbers in the hall. CoBot can overcome its limitations by asking humans for help.

Other robot architectures allow the robot to ask for help to overcome limitations, most notably social learning (Asoh et al. 1997) collaborative control (Fong, Thorpe, and Baur 2003; Breummer et al. 2005), and sliding autonomy (Heger et al. 2005)). However, in these models, the human is not receiving any help or benefit from the robot in exchange for providing help. CoBot, instead, requests help from the visitor which, in return, helps it more successfully complete it's tasks (Rosenthal, Biswas, and Veloso 2010). For example, in order to navigate the visitor to meetings efficiently, the robot requires accurate localization. If the robot is uncertain of its location, it may ask "Can you point to where we are on this map?" which requires the visitor to stop and read the office numbers (Figure 2). Once the visitor responds, the robot can then continue escorting the visitor to her next meeting. While currently, the visitor responds to the help requests on

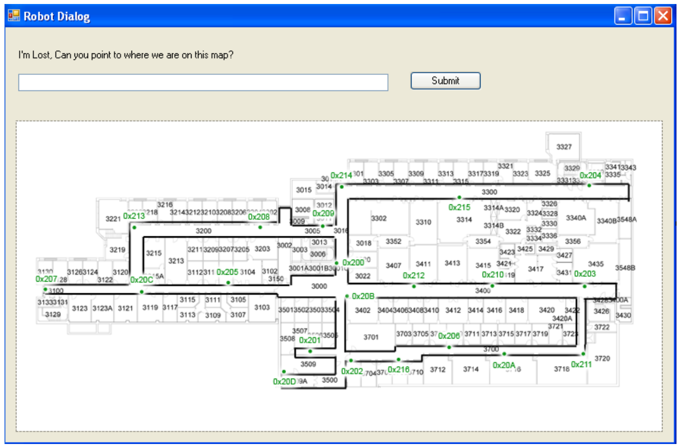

Figure 2: The visitor clicks on the map on CoBot's screen to indicate their location.

a user interface on the laptop, this interface is not necessarily the most understandable for visitors who do not know the map of the building and could be extended to a speech interface as well.

In order to take full advantage of the human help, we would like the visitor to answer the robot's questions as accurately as possible. However, the visitor may not always be accurate if she does not understand the question or does not have enough contextual information to know what the robot is referring to. Recently, much research has focused both on modifying interfaces to make it easier for human helpers to understand what the robot is asking (*e.g.,* (Argall, Browning, and Veloso 2007; Shiomi et al. 2008)) and to make the robots more robust to inaccurate answers (*e.g.,* (Chernova and Veloso 2008)). Next, we describe our work towards obtaining more accurate responses to the robot's help requests by altering the robot's dialog.

Asking For Help

In order to understand how the content of CoBot's questions affects the accuracy of the visitor's responses, we draw from both the dialog literature and the human-computer interaction literature. First, perhaps most widely cited is the work on *grounding* humans in a frame of reference (Clark, Schreuder, and Buttrick 1983). Because a human likely does not know about a robot's sensors, a robot that can explain the data that it is using to make a prediction may help a human answer the question more accurately (Steels 2003; Clark 2008). Additionally, it is widely known that humans are better at recognizing an answer rather than recalling it (Tulving and Wiseman 1975). Fong et al's robot, for example, provided usual or predicted responses to its own questions when asking a novice human for navigation feedback (Fong, Thorpe, and Baur 2003). However, humans, are less likely to ask for help and respond to help requests if they feel that the question should be easy to solve without their help (Depaulo and Fisher 1980). By revealing uncertainty, or the probability that the prediction is incorrect, people can prioritize the need to respond to a robot's request for help (Shiomi et al. 2008). Finally, just as humans ask for reasons why they receive certain responses, forms of corrective feedback have been shown to aid machine learners increase classification accuracy without increasing the cognitive overhead of the responder (Stumpf et al. 2007).

We performed two experiments to understand whether each of these dialog features (grounded context, uncertainty, prediction, and asking why) would increase the accuracy of human responses to a robot's questions in 1) a shape recognition task on a Wizard of Oz'd robot and 2) the localization task on CoBot. In the first task, participants were given the task to build towers out of blocks and were told that an observing robot would sometimes ask them to identify the shape (cube, cylinder, etc) (Rosenthal, Dey, and Veloso 2009). In the localization task, visitors were given a tour of our building and were asked to click on a map to identify their location (Rosenthal and Veloso 2010). In both studies, we varied whether the robot provided each of our four kinds of information (more details in (Rosenthal, Dey, and Veloso 2009)). Participants were given different subsets of 1) a description of the question-relevant context (the block colors near the block in question or the hallway the robot was located in), 2) the robot's prediction of the answer (the block shape or a dot on the map indicating location), 3) its uncertainty in that prediction, and 4) an additional question asking humans to describe their answer.

The results of our initial shape recognition task show that providing users all four of the types of information together improved the accuracy of their responses to the robot the most. While we realize providing all of this information is not always time effective or possible given different robot limitations, we are able to provide guidelines for increasing accuracy when accuracy is very important. We validated this combination of information in the second experiment in which actual building visitors responded to CoBot's question on the same map interface ("Can you point to where we are on this map of the building") while taking a tour of the building. Participants in this study were randomly assigned to one of five information conditions: 1) the question without any information, 2) uncertainty and context, 3) uncertainty and prediction, 4) uncertainty, context, and prediction, and 5) uncertainty, context, prediction, and extra question. The experimenter remote-controlled the robot to each tour location in the building and triggered questions at 13 known locations. After participants completed the 15-minute tour and responded to all localization questions, they were given a survey about their experiences with the robot.

Our results validate our shape recognition study findings and show that providing all four types of information to humans can significantly increase the accuracy of the responses those humans give. We find that the context and prediction helped narrow the focus of the participants' responses to questions. Participants who received uncertainty information responded that they felt the robot needed their help and that the additional question "why" was very difficult to answer (although it was shown to have increased participants' accuracy). Using the results of the localization study, we have also been able to show that CoBot can navigate more efficiently with fewer backtracks when it receives accurate responses compared to autonomous navigation (Rosenthal, Biswas, and Veloso 2010). Next, we outline our plans to ensure that asking for help and providing notifications is not repetitious.

Addressing Repetitive Dialog

During an all-day interaction, there may be many times when the robot must ask the same question and could notify the visitor of the same interesting place. We propose employing two different techniques to keep the interaction more engaging by reducing the repetitiveness. First, rather than repeating the same notifications about the same locations through the day, CoBot gives more detailed information each time it passes, assuming that the visitor is interested in and available to hear that information. Second, our future work includes probabilistically choosing an utterance from a set of synonyms so that similar notifications do not sound repetitious.

Adding Information

Because it is likely that the visitor will pass through the same areas multiple times, we would like to ensure that CoBot does not repeat the same information about the same location twice. However, we also would like to ensure that CoBot can continue sharing information through the day to keep the visitor engaged. Towards these goals, CoBot shares more information about locations as it passes them multiple times. For example, the first time that CoBot passes a robot lab, it provides brief information about the faculty who runs the lab and the types of robots available. The second time, it provides more detailed information on the motivation for their work. The third time, it offers examples of specific projects or results that the lab has shown recently.

In order for the dialog manager to provide increasingly detailed information, it must maintain a count of how many times the robot has passed each location. Currently, CoBot's notifications are ordered in terms of detail for each known location in the building. This will become cumbersome as the number of CoBot's known locations increases. In the future, we would like to parse lab websites for up-to-date and dynamic information about more locations.

Using Synonyms

In order to reduce the repetition in notifications, we propose probabilistically choosing the wording of notifications from a set of synonyms like the CMCast system (Veloso et al. 2008). The CMCast robot commentators watched the RoboCup soccer games and explained the game to the spectators. Because the same plays occurred multiple times as well as multiple goals, the commentators varied their speech through the game, assigning weights to their utterances based when it was spoken last and probabilistically choosing the utterance to speak next in proportion to the weights. For example, if there are two different ways to announce a goal - "GOAL!" and "The fourth goal for the blue team" - the utterance that was spoken most recently would have lower weight and would be less likely to be spoken at the next goal.

There are many ways for CoBot to respond to a request, offer help, and ask for help. We will vary the utterances to ensure that the dialog does not become boring and predicable through the day by similarly weighing the responses by both the frequency and recency of the utterance.

Example Dialog

We present an illustrative trace of CoBot's core dialog capabilities. This trace is just one of many such possible sequences of events that may occur depending on the environment and the visitor's needs.

CoBot and the visitor start in the initial location Room 3716, and both know of the schedule below. Alice spends 5 minutes training the speech recognizer for her voice. CoBot explains that it has received her schedule and will help her get to each meeting.

```
Visitor:  Alice
Interests:  Robotics
Schedule:
9:00-10:00 AM - John Smith, Room 3123
10:00-11:00 AM - Mary Jones, Room 3714
11:00-12:00 PM - Mike Adams, Room 3456
```

`8:50 AM - CoBot asks if Alice wants a drink.`
Task Offering - Alice replies that she would like coffee. CoBot's adds the subtask of getting coffee when it is next convenient (either on the way to the meeting or after dropping off Alice at the meeting). When Alice is ready, the robot begins escorting her to Room 3123 (Figure 3a).

`8:55 AM - CoBot asks Alice for localization help when it gets lost.`
Help Request - As CoBot navigates to Room 3123, it becomes uncertain of its location. It stops and asks Alice to click on a map to indicate their location. When Alice responds, CoBot continues navigating.

`9:00 AM - CoBot and Alice arrive at Room 3123. CoBot leaves to retrieve coffee while Alice is engaged in the meeting.`
Help Request - CoBot navigates to an administrator's office

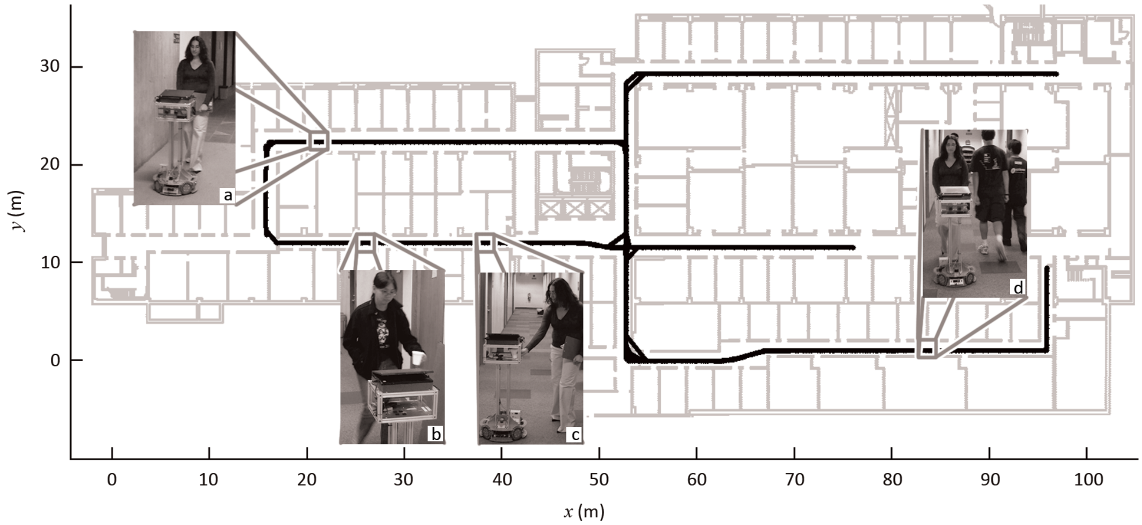

Figure 3: Trace of the path traversed by the robot while following the schedule. Snapshots: a) CoBots leads Alice to her first meeting b) CoBot requests assistance with getting a cup of coffee c) Alice graciously accepts her cup of coffee d) CoBot notifies Alice of a lab of interest.

to ask for help in getting coffee because it cannot pour the coffee itself (Figure 3b).

`9:10 AM - CoBot returns with the coffee.`
Notification - CoBot returns with the coffee that the administrator helped pour (Figure 3c).
Help Request - Because it recognizes that its battery power is low, CoBot also asks "Can you please plug me in until we're ready to go?".

`9:55 AM - CoBot asks Alice if she's ready to go to her next meeting.`
Task Offering - Shortly before the next meeting begins, CoBot notifies Alice that her next meeting will start soon. When Alice and the host finish talking, CoBot begins navigating to Room 3714.

`10:00 AM - Alice asks for more information about her next meeting host, Mary Jones.`
Task Request - Alice requests more information about Mary, her next meeting host. CoBot stops navigating and displays the host's website for her to read. When Alice is done, CoBot starts moving again.

`10:05 AM - CoBot notifies Alice of about a robot lab.`
Notification - As CoBot passes a robot lab that Alice may be interested in, it gives a short description of the current projects there (Figure 3d). CoBot and Alice arrive in time for the meeting with Mary in Room 3456.

`11:00 AM - CoBot emails Mike that Alice is running late.`
Task Offering - At 11:00, Alice is still talking to Mary. CoBot emails Mike that Alice is running late.

`11:05 AM - CoBot provides more information about the robot lab.`
Notification - As CoBot passes the same robot lab, it provides additional information about the motivation for the research.

`11:05 PM - Alice requests a taxi to the airport at the end of the day.`
Task Request, Help Request - Alice requests a taxi to the airport. CoBot emails an administrator to ask for help with that task. When it gets a confirmation, it adds the taxi location to the schedule and escorts Alice to that location when the meeting with Mike is over.

Conclusion

As robots start to be used in longer-term tasks and interactions, it is imperative for them to be able to converse with humans in their environments. Our robot, CoBot, is designed to participate in long-term interactions with a visitor, escorting her to meetings around a building. CoBot engages in mixed-initiative task-oriented dialog with the visitor to ensure that the visitor receives all of the amenities, such as coffee, she wants. The visitor can make task-related requests and CoBot maintains state about the visitor so that it can proactively offer amenities at appropriate times. Because CoBot is near the visitor for a majority of the time, it can take advantage of the visitor's strengths in localization as well as physical manipulation by asking for the visitor's help to perform these tasks. In particular, we have shown that CoBot can change the way it asks for localization help, by including more contextual information in the question, in order to receive more accurate responses from the visitor.

Beyond improving the visitor's accuracy, we would like the visitor to receive interesting and non-repetitive information and notifications from CoBot through the day. CoBot provides increasingly detailed information about locations of interest as it passes the same locations over and over and we have future plans to vary other utterances in the dialog manager to prevent repetition. Our future work includes deploying CoBot with visitors to test the usability of its mixed-initiative dialog in all-day interactions and its overall ability to satisfy the Visitor-Companion Task.

Acknowledgments

We would like to thank Anind Dey for invaluable discussions and input on human-robot interaction and Joydeep Biswas for CoBot's localization and navigation algorithms and implementation. We further thank the participants who generously provided their time in our experiments.

This research was partially supported by a Fellowship from the National Science Foundation and the Computational Thinking Center at Carnegie Mellon. The views and conclusions contained in this document are those of the authors only.

References

Argall, B.; Browning, B.; and Veloso, M. 2007. Learning by demonstration with critique from a human teacher. In *International Conference on Human Robot Interaction*, 57–64.

Asoh, H.; Hayamizu, S.; Hara, I.; Motomura, Y.; Akaho, S.; and Matsui, T. 1997. Socially embedded learning of the office-conversant mobile robot jijo-2. In *International Joint Conference on Artificial Intelligence*, volume 15, 880–887.

Biswas, J., and Veloso, M. 2010. Wifi localization and navigation for autonomous indoor mobile robots. In *IEEE International Conference on Robotics and Automation*.

Bohus, D., and Horvitz, E. 2009. Dialog in the open world: platform and applications. In *International Conference on Multimodal Interfaces*, 31–38.

Breummer, D. J.; Few, D. A.; Boring, R. L.; Marble, J. L.; Walton, M. C.; and Nielsen., C. W. 2005. Shared understanding for collaborative control. *IEEE Transactions on Systems, Man, and Cybernetics* 35(4):494–504.

Chernova, S., and Veloso, M. 2008. Learning equivalent action choices from demonstration. In *International Conference on Intelligent Robots and Systems*, 1216–1221.

Clark, H.; Schreuder, R.; and Buttrick, S. 1983. Common ground and the understanding of demstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22:245–258.

Clark, H. 2008. Talking as if. In *International Conference on Human Robot Interaction*, 393–394.

Depaulo, B., and Fisher, J. 1980. The costs of asking for help. *Basic and Applied Social Psychology* 1:23–35.

Fong, T. W.; Thorpe, C.; and Baur, C. 2003. Robot, asker of questions. In *Robotics and Autonomous Systems*, volume 42, No. 3-4, 235–243.

Gockley, R., et al. 2005. Designing robots for long-term social interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*, 1338–1343.

Heger, F.; Hiatt, L.; Sellner, B.; Simmons, R.; and Singh, S. 2005. Results in sliding autonomy for multi-robot spatial assembly. In *International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 5–8.

Horvitz, E. 1999. Principles of mixed-initiative user interfaces. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, 159–166. New York, NY, USA: ACM.

Kanda, T.; Shiomi, M.; Miyashita, Z.; Ishiguro, H.; and Hagita, N. 2009. An affective guide robot in a shopping mall. In *International Conference on Human Robot Interaction*, 173–180. ACM.

Lee, M. K.; Forlizzi, J.; Rybski, P. E.; Crabbe, F.; Chung, W.; Finkle, J.; Glaser, E.; and Kiesler, S. 2009. The snackbot: documenting the design of a robot for long-term human-robot interaction. In *International Conference on Human Robot Interaction*, 7–14.

Pineau, J.; Montemerlo, M.; Pollack, M.; Roy, N.; and Thrun, S. 2003. Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems* 42(3-4):271–281.

Rich, C., and Sidner, C. L. 1998. Collagen: A collaboration manager for software interface agents. *User Modeling and User-Adapted Interaction* 8(3-4):315–350.

Rosenthal, S., and Veloso, M. 2010. Using symbiotic relationships with humans to help robots overcome limitations. In *Workshop for Collaborative Human/AI Control for Interactive Experiences*.

Rosenthal, S.; Biswas, J.; and Veloso, M. 2010. An effective personal mobile robot agent through a symbiotic human-robot interaction. In *International Joint Conference on Autonomous Agents and Multiagent Systems*.

Rosenthal, S.; Dey, A. K.; and Veloso, M. 2009. How robots' questions affect the accuracy of the human responses. In *The International Symposium on Robot - Human Interactive Communication*, 1137–1142.

Shiomi, M.; Sakamoto, D.; Takayuki, K.; Ishi, C. T.; Ishiguro, H.; and Hagita, N. 2008. A semi-autonomous communication robot: a field trial at a train station. In *International Conference on Human Robot Interaction*, 303–310.

Steels, L. 2003. Evolving grounded communication for robots. *Trends in Cognitive Sciences* 7(7):308–312.

Stumpf, S.; Rajaram, V.; Li, L.; Burnett, M.; and Dietterich, T. 2007. Toward harnessing user feedback for machine learning. In *International Conference on Intelligent User Interfaces*, 82–91.

Thrun, S., et al. 2000. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *The International Journal of Robotics Research* 19(11):972–999.

Tulving, E., and Wiseman, S. 1975. Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society* 6:79–82.

Veloso, M.; Armstrong-Crews, N.; Chernova, S.; Crawford, E.; McMillen, C.; Roth, M.; Vail, D. L.; and Zickler, S. 2008. A team of humanoid game commentators. *International Journal of Humanoid Robotics* 5(3):457–480.