

Effects of Adaptive Robot Dialogue on Information Exchange and Social Relations

Cristen Torrey¹, Aaron Powers¹, Matthew Marge², Susan R. Fussell¹, Sara Kiesler¹

¹Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15217

ctorrey, apowers, sfussell, kiesler@cs.cmu.edu

²Department of Computer Science
Stony Brook University
(SUNY)
Stony Brook, NY 11794-4400
mmarge@ic.sunysb.edu

ABSTRACT

Human-robot interaction could be improved by designing robots that engage in adaptive dialogue with users. An adaptive robot could estimate the information needs of individuals and change its dialogue to suit these needs. We test the value of adaptive robot dialogue by experimentally comparing the effects of adaptation versus no adaptation on information exchange and social relations. In Experiment 1, a robot chef adapted to novices by providing detailed explanations of cooking tools; doing so improved information exchange for novice participants but did not influence experts. Experiment 2 added incentives for speed and accuracy and replicated the results from Experiment 1 with respect to information exchange. When the robot's dialogue was adapted for expert knowledge (names of tools rather than explanations), expert participants found the robot to be more effective, more authoritative, and less patronizing. This work suggests adaptation in human-robot interaction has consequences for both task performance and social cohesion. It also suggests that people may be more sensitive to social relations with robots when under task or time pressure.

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, Natural language, Theory and methods*.

General Terms

Design, Experimentation, Human Factors, Performance, Theory.

Keywords

Human-robot interaction, social robots, human-robot communication, common ground, collaboration, perspective taking, adaptive dialogue.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

HRI'06, March 2–4, 2006, Salt Lake City, Utah, USA.
Copyright 2006 ACM 1-59593-294-1/06/0003...\$5.00.

1. INTRODUCTION

In this paper we explore how social robots might use adaptive dialogue to advise, instruct, guide, test, or interview a varied group of individuals. In these roles, the robot may need to help people understand instructions and identify objects, locations, or tools. We address the possible benefits of a robot adapting to individual differences in people's knowledge and the possible costs of not doing so. The goal of our research is to improve our understanding of how best to achieve effective natural language communication with robots.



Figure 1. Pearl as a robot chef.

For numerous social roles, natural language dialogue seems particularly appropriate for robots [22]. For example, robot receptionists [13] and museum guides [24] can respond to questions. A tutor robot speaks English with Japanese children [17]. Robots using natural language can certainly be engaging [29], thus the number of robots that respond in natural language is growing. In addition, there are an increasing number of robots that respond to users via tone of voice, gaze, and gesture [3, 4]. The robot in Figure 1 was initially designed to interact with elders [25]. In the current research, the same robot serves as a chef robot, instructing novice and expert cooks with a male voice and responding to users' typed input.

Interactions with members of the public may particularly benefit from the use of adaptive dialogue. Robots designed as advisors or guides for public settings interact with a population diverse in its interests, background, and information requirements. An out-of-town visitor conversing with a robot receptionist has different information needs than an employee. To a visitor seeking directions to someone's office, the robot might need to say, "When you get to the red brick building in about 2 blocks, make a

left. The building you want is yellow and has a sign in the front, Blaylock A. Doctor Smith is on the second floor. You can take the elevator.” To an employee (for whom Blaylock A is a well-known landmark), the robot might say “Dr. Smith is in Blaylock A, second floor.”

Little is known, thus far, about adaptivity in human-robot dialogue. The human-robot interaction community needs to learn if adaptivity will be worth the difficulty before creating robots that can identify, assess, and respond appropriately to individual differences in people. Generally, social robots respond in scripted ways that do not account for individual differences. For instance, Valerie the roboceptionist allows people to swipe an ID card to identify themselves, but this awareness has not yet been used to modify the robot’s dialogue for different types of people [13]. User modeling has been of interest to the dialogue systems community for some time. Some systems in the domain of travel scheduling do attempt to account for the knowledge and preferences of individual users [19, 21], although the adaptation in these systems is not embodied. A related development is the consideration of the individual’s perspective in physical space. Using a model of the listener’s spatial perspective, the robot can refer to objects in the shared physical space from the listener’s point of view [30].

In the remainder of this paper, we first outline the theoretical background and hypotheses guiding our work on robot adaptivity. Our domain of interest is robots that can modify their language to meet listeners’ information needs. Then, we present two human-robot interaction experiments in which the robot responds either in a manner appropriate for experts or in a manner appropriate for novices. We conclude with a discussion of the limitations, significance, and future directions of this work.

2. THEORY AND HYPOTHESES

We derive our theoretical framework from the literature on common ground and the grounding process that unfolds in conversation [7, 27]. People draw upon the knowledge and beliefs they share with their listeners when they formulate their messages. To identify this common ground, they make use of cues to listeners’ attributes such as their age, gender or group memberships [9]. In addition, speakers use listener feedback to refine their models of listeners’ expertise and adapt subsequent communications to meet these needs [16]. Communications designed specifically for a listener are understood more easily and result in more efficient communication than messages created for someone else or a generic listener [10, 28].

Clark and Wilkes-Gibbs [6] have proposed the concept of “least collaborative effort” to explain why messages that are adaptive to a listener’s level of expertise are more successful. With appropriate messages, listeners can simply say “ok” or otherwise indicate that they understand. In contrast, messages that are inappropriately adapted to listeners’ expertise will require more overall effort by both parties. If the message is too detailed, as would be the case if directions for an out-of-towner were given to a local resident, the speaker has put forth more effort than necessary. If the message is not detailed enough, as would be the case if directions for a local resident were given to an out-of-towner, subsequent clarifying discussion will be necessary.

Adapting to one’s audience not only improves communication efficiency, but also helps maintain positive affect between speakers and listeners. When too little information is provided, listeners may interpret the sparse information as a sign that the speaker has no concern for their needs. Similarly, when too much information is provided, listeners may feel insulted. In general, people are motivated to maintain each other’s “face” or positive impression of themselves [14]. One way in which speakers do so is by providing listeners with the right amount of information for their needs. Communications that threaten face can lead to negative evaluations of a speaker [15]. Appropriate adaptation has further been shown to facilitate social coordination and have other broad-reaching benefits for interaction [11, 12].

We can apply the theories of conversational grounding and literature on face saving to human-robot interaction in the following way. First, we assume that a robot can assess certain individual differences in users’ needs for information. For example, if the robot can identify individuals, it can distinguish between employees and first-time visitors. The robot can use social network data to estimate associations, and thus domains of knowledge. The robot could also use physical cues to people’s gender, nationality, or age to assess their expertise. Or, the robot might use conversation to assess the knowledge of others. Second, we assume the robot can vary its dialogue to suit its estimate of the knowledge of others, thereby facilitating the grounding process. Thus, when a robot gives directions to a person knowledgeable about the local area, it can assume more common ground than it does when it gives directions to a stranger. In the former case, the robot can use efficient terminology, such as names of landmarks, whereas in the latter case, it will need to provide more detailed explanation. Following the principle of least collaborative effort, we argue that the robot’s adaptation to the information needs of individuals will increase the efficiency of information exchange. From the literature on face saving, we argue, further, that a robot that adapts to the knowledge of individuals will help them maintain face and improve their evaluation of the robot, as they will perceive the robot as understanding their needs and caring about them.

From the above arguments, we predict that information exchange between a robot and novice versus a robot and expert will be differentially improved by an adaptive robot. The adaptive robot will provide more explanation and detailed description to novices than to experts. Novices will gain the information they need to succeed in their task when the robot adapts to their greater need for information. They will be less likely to succeed if the robot is not adaptive to their information needs. By contrast, experts’ understanding and performance will not be affected. An adaptive robot should provide them with less explanation because they do not need it, but more explanation will not hurt.

Hypothesis 1: Information exchange. The performance of novices will benefit from additional descriptive information, whereas the performance of experts will not.

We make a different prediction for the social relations between a robot and novice versus a robot and expert. That is, the attitudes of both novices and experts toward the robot will be influenced by whether or not it is adaptive. Novices should appreciate a robot that adapts to their greater information needs, and experts should appreciate a robot that adapts to their lesser information needs. If the robot treats experts as though they are novices, the extra

explanation does not hurt their task performance, but it does indicate a disregard for their level of expertise. Thus, experts should not like the nonadaptive robot.

Hypothesis 2: Social Relations. Novices will like the robot more when offered additional descriptive information, whereas experts will like the robot less.

3. EMPIRICAL STUDIES

We examined the value of adaptive dialogue by comparing the outcomes of receiving adaptive dialogue with the outcomes of receiving nonadaptive dialogue. The two experiments engaged participants in a cooking-related task aided by a robot with expertise in cooking.

To identify adaptive and nonadaptive descriptive language for individuals with different expertise, we developed a short test to categorize expert and novice cooks. We placed novices and experts in two experimental conditions and observed their interactions with the chef robot. The robot asked participants to identify pictures of cooking tools used in making crême brûlée, a gourmet dessert of custard topped with caramelized sugar. In one condition, the robot simply named the cooking tools. This dialogue was adapted for expert cooks. In the other condition, the robot described and explained the function of each tool. This dialogue was adapted for novices.

We measured two kinds of outcomes. The research on human-human communication cited above suggests that adaptation affects information exchange and social relations. We thus measured participants' understanding of the robot and their ability to choose the right cooking tools, and we measured participants' regard for the robot and their attributions of its personality traits.

3.1 Identifying Experts & Novices

We developed a simple test to measure participants' cooking expertise. Sixteen pilot participants identified ten cooking tools from among a set of possible choices (see example in Figure 2). These participants also answered eight questions in which they matched definitions of cooking methods with the names of these methods. Pilot participants who could match the cooking methods with their definition were also able to correctly identify the cooking tools ($r = .71$, $F [1, 15] = 15.4$, $p = .001$).



Figure 2. Screen display for cooking tool selection task.

Potential participants for the subsequent experiments were asked to complete a short pretest, matching the cooking methods with their definitions. Participants who scored 100% on this pretest were classified as experts; they are likely to be able to identify all of the cooking tools by name as well. Those who scored less than 50% were classified as novices.

Although it would have been possible for the robot to identify participants' expertise at the time of the experiment through an interview (asking them the same questions that were on the pretest), we chose to differentiate between novices and experts by giving the test before the experiment. Our purpose was to remove potential participants with ambiguous levels of expertise, that is, those who scored better than 50% but less than 100% on the test. In doing so, we could efficiently test our hypotheses.

3.2 Experiment 1

Experiment 1 tested the effects of dialogue that was matched to individuals' expertise compared to dialogue that was mismatched. We predicted (Hypothesis 1) that novices who received mismatched dialogue would experience a negative impact on the quality of information exchange. We also predicted (Hypothesis 2) that both novices and experts who received mismatched dialogue would experience a negative impact on the social relation dimension.

To investigate our hypotheses, we used the cooking tool selection questions exactly as they appeared in the pilot testing. The participants' goal was to select ten cooking tools needed to make a crême brûlée dessert. Participants selected the tool by clicking on the correct picture on a computer monitor. Each of the ten tools was displayed separately alongside five incorrect tools. The robot conversationally led the participant through the task, requesting each of the tools in turn, and answering participants' questions. Participants could ask the robot as many questions as they wished.

3.2.1 Method

The experiment was a 2X2 (expertise X dialogue) between-subjects design. We varied expertise, as previously described, by administering an online test prior to participation and selecting novices and experts for comparison. We created two dialogue conditions. In the first condition, the "Names Only" condition, the robot directed the participant to the tool by identifying the tool by name. This condition was hypothesized to be more suitable for experts. In the second condition, the "Names Plus Description" condition, the robot named the tool and further described it in several sentences. This condition was hypothesized to be more suitable for novices with little knowledge about the proper names of cooking tools (see Table 1 for example description).

Table 1. Example directions for finding the paring knife.

Condition	Robot Dialogue
Names Only	Next you want a sharp paring knife. Find the paring knife.
Names Plus Description	Next you want a sharp paring knife. Find the paring knife. It's usually the smallest knife in a set. It has a short, pointed blade that is good for peeling fruit. The blade is smooth, not jagged.

3.2.1.1 Participants

Forty-nine students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University. They were each paid \$10 for their participation in this experiment.

3.2.1.2 Robot

The robot used for this experiment was originally designed to interact with people in a nursing home [25]. In this experiment, the robot was stationary and was dressed to appear like a cooking expert. The robot wore a white chef's hat and apron and spoke with a male voice. The robot opened its eyes at the start of the experiment and closed them at the end. While speaking, the robot's lips moved in synchrony with its words. The robot's face is articulated and is capable of a range of expressions, but the full range of expression was not utilized in this experiment. By limiting the robot's movement, we were better able to isolate the effect of the robot's language use.

3.2.1.3 Procedure

When participants arrived at the experimental lab, the experimenter told the participant that the robot had been given "specific expertise" in cooking, and that "the robot will be talking to you about the tools needed to make a cr me br l e dessert."

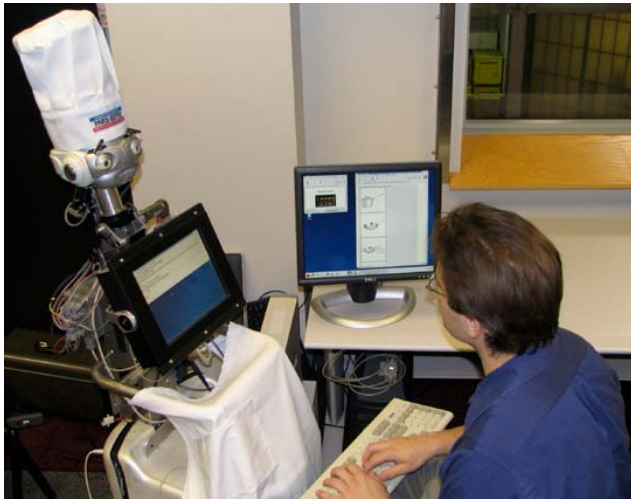


Figure 3. The experimental set-up.

The robot spoke aloud and also displayed its messages on a display on the robot's chest. The robot used Cepstral's Theta [18] for speech synthesis, and its lips moved as it spoke. The text also showed on the screen, as in Instant Messenger interfaces. The interface was identical to the interface in [26] except that the dialogue technology was improved further, as discussed in the next section of this paper.

Participants interacted with the robot by typing into the same Instant Messaging interface. We used a robot without speech recognition because of current limitations in speech understanding across individuals when the dialogue is complex, as in the current case.

In the course of the dialogue, the robot prompted the participant to find cooking tools, e.g., "Find the picture of the saucepan." The tools were shown on a nearby computer (see Figure 3). If the

participants knew which tool was correct, they clicked the correct image and told the robot that they found the right tool. If the participant did not recognize the name of the cooking tool, they could ask the robot questions about the tool, using the IM interface (some participants spoke out loud as well). Most of participants' questions were about tool properties like shape ("does it have a round bottom"), color ("what color is it"), and usage ("what is it for"). The robot was programmed to respond to most of these inquiries.

All the participants' responses to the robot were logged. After conversing with the robot, the participant completed a survey about their perceptions of the robot and their conversational interaction.

3.2.1.4 Dialogue Technology

The robot interpreted and responded to participant input using a customized variant of Artificial Intelligence Mark-up Language (AIML) [31], a publicly available pattern-matching text processor. In previous experiments, we found that the existing implementation of AIML could not respond well to participants' questions, in part because it could not make use of dialogue context. To gain more control over the flow of the dialogue, we wrapped another technology layer around AIML and made significant changes to how AIML is processed. These modifications greatly improved the robot's ability to understand and respond intelligently.

We modified the dialogue empirically through iterative pilot testing. For instance, in the course of the experiment, participants had to tell the robot that they found the tool. Participants expressed this in different ways, so we compiled a list of over 50 phrases for confirming the user had made a choice, such as "I made a guess," and "I actually knew that one." In the second experiment, the robot misunderstood this type of expression only once out of 480 confirmation-related responses.

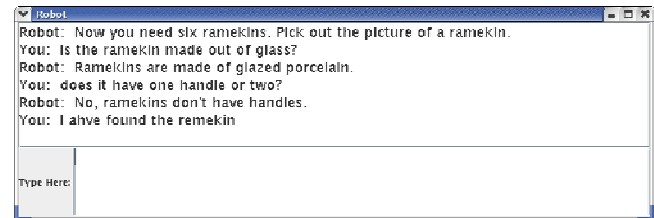


Figure 4. The robot's responses were automated with a customized version of AIML.

Participants also asked different kinds of questions about cooking tools, such as "is the ramekin made out of glass," or "does it have one handle or two?" (See Figure 4 for an example dialogue.) To meet this need, we created an AIML-based database of cooking tools, tool properties, and answers to questions about tool properties. We ran pilot experiments and examined participants' questions, focusing on questions that the robot did not answer correctly. After each successive pilot test, and again following Experiment 1, we revised the knowledge base and iteratively improved the coverage of the robot's responses.

We also improved the search algorithms AIML uses to find matching responses. We replaced AIML's depth-first search with an A* search, and we added priorities and a system for finding the best match to a question (formerly, AIML stopped searching on the first find). These and other changes drastically reduced the

robot’s errors. In Experiment 1, when the robot answered a participant’s questions, the robot’s error rate due to AI failure was only 15%, and most of that (14.5%) was because the robot’s knowledge base did not contain that information about the object. When the robot did not know the answer to a question, it either said “I don’t know” or it gave a little more information about the object (“It [the saucepan] is medium-size, maybe two quarts, and has high, straight sides”).

Participants in the pilot studies for this and previous experiments sometimes made spelling and grammatical errors, for example, “ahve” instead of “have” (see last line, Figure 4). To interpret these responses, we used the Aspell spell checker [1] to find spelling errors and correct them automatically in the robot’s interpreter. When the AI did not understand a word, it tried many alternative spellings and used the best match.

3.2.1.5 Measures

We measured interaction with the robot on two dimensions, information exchange and social relations. In the course of telling the participant about making crême brûlée, the robot asked the participant to identify ten cooking tools. We measured information exchange as the number of questions participants asked about the tools. A greater frequency indicates the participant did not know which tool was correct and needed more information. We also measured the accuracy of their performance as the number of tools they got correct. We did not expect large differences in performance, because the participants could keep asking the robot questions until they thought they understood which cooking tool was correct. We also measured the time each participant spent on the task and the number of misunderstandings with the robot.

We assessed participants’ social relations with the robot through self-report items on a questionnaire. Participants completed this questionnaire following their interaction with the robot. The questionnaire covered three general areas of interest: participants’ impressions of the robot’s personality and intellectual characteristics (authority, sociability, intelligence), participants’ evaluation of the quality of the communication (effectiveness, responsiveness, control), and participants’ evaluations of the task (enjoyability, ease).

Table 2. Scale reliabilities for measures of social relations.

Scale (Cronbach’s Alpha *)	Sample Item
Robot Authority (0.72)	Expert/Inexpert
Robot Responsiveness (0.76)	My partner can adapt to changing situations.
Conversational Control (0.86)	My partner was more active in the conversation than I was.
Conversational Effectiveness (0.90)	I found the conversation to be very useful and helpful.
Task Difficulty (0.77)	This task was difficult.
Task Enjoyability (0.75)	I enjoyed participating in this task.

* Cronbach’s Alpha is a measure of the reliability of the scale as a whole. Alpha ranges from zero to 1.0 (highest).

To assess users’ perception of the robot’s authority [20], and sociability and intelligence [32], we used existing scales from the social psychology literature. We used these scales in their entirety. We also selected items from a published (but lengthy) communicative effectiveness scale [5] and from a communicative competence scale [33]. We also created scales for task enjoyability and task difficulty. We conducted factor analysis on the scales after collecting data in Experiment 1 to confirm that the scales were appropriate for evaluating a robot. The reliabilities for the scales are shown in Table 2.

3.2.2 Experiment 1 Results & Discussion

The robot asked participants to identify 10 cooking tools. We considered the participants’ number of questions as the measure of information exchange. The number of questions reflects the amount of uncertainty the participant has about which cooking tool is correct and is related to the amount of effort exerted by the participant in communicating with the robot. As pictured in Figure 5, there was a significant interaction between expertise and the dialogue condition ($F [3, 48] = 9.99, p < .01$). Novices were negatively impacted by the absence of additional detail in the Names Only condition. Experts were not harmed by the additional detail the robot gave them in the Names Plus Description condition. We thus find support for our first hypothesis. Without an appropriate amount of detail, novices had to work harder to communicate with the robot and get the information they required.

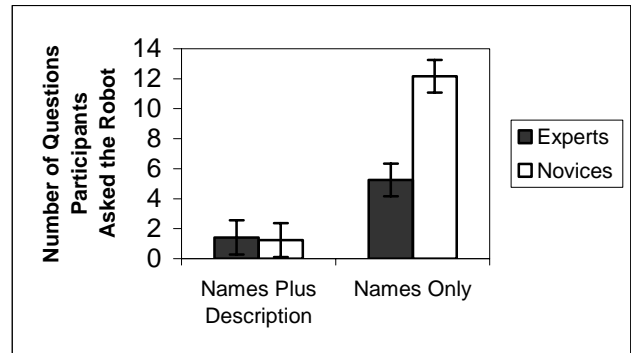


Figure 5. Novice users were affected disproportionately by the robot’s lack of description.

Our second hypothesis predicted that a mismatch between expertise and dialogue condition would have consequences for social relations. We checked our measures of social relations via the questionnaire data, but there were no significant interactions. Because the length of the interaction was short and there were no incentives for finishing quickly, we considered the possibility that participants may not have felt strongly about the social consequences of too much or too little information even if the communication was ineffective. To make the task more compelling, we decided to add a monetary incentive for accuracy and a monetary incentive for speed to the experimental design. In the second experiment, our purpose was to create an experimental set-up in which the robot would really be contributing to their success or failure.

In examining novice task performance, we found that novices made frequent errors on several of the cooking tools even when given an explicit description by the robot. In Experiment 2, we

described these tools in even greater detail in the Names Plus Description condition. Our intention with this additional detail was to ensure success for novice users. For example, the new description of the paring knife added detail that the blade was not curved and emphasized that the blade was short.

3.3 Experiment 2

Our first hypothesis was confirmed in Experiment 1. Inappropriate adaptation does affect the amount of effort novices must exert in the process of information exchange. However, we did not see any effect of the interaction on social relations. To further investigate social relations in human-robot interaction, we added the element of time pressure to the conversation. We predicted that under pressure, giving experts information about which they are already familiar will strain the social relationship and will result in experts evaluating the robot more negatively.

3.3.1 Method

The design of Experiment 2 was the same as Experiment 1. The procedure was the same with the exception that we added incentives for speed and accuracy and, in the Names Plus Description condition, increased the detail of explanation of the three tools that novices were particularly likely to choose incorrectly in Experiment 1.

3.3.1.1 Participants

Forty-eight students and staff members with no prior participation in our experiments were recruited from Carnegie Mellon University. They were each paid \$8 plus possible bonuses up to \$15 for participation in the experiment.

3.3.1.2 Procedure

To create time pressure during the experiment, participants were informed in the written instructions that if they finished the task quickly, they would receive an additional \$1 for every minute under the average participant time. The experimenter answered any questions about the experiment and started a timer when the participant typed "hello" to the robot to begin the task. We also had an incentive for accuracy. If participants correctly identified all ten items, they would receive an additional \$3 in payment. We displayed a running timer on the monitor where the participants

were selecting the cooking tools. When they began conversing with the robot, we started the timer, and it was visible the entire time they worked at the computer.

3.3.1.3 Measures

In addition to the measures gathered in the first experiment, we added eight questions to the post-experiment questionnaire. We predicted these scales would load on two social relations factors, patronizing communication and content appropriateness. The questionnaire data from Experiment 2 supports these concepts as scales (see Table 3). We added these to further explore the negative social consequences of nonadaptivity on experts.

Table 3. Scale reliabilities for additional social measures.

Scale (Cronbach's Alpha)	Sample Item
Patronizing (0.90)	My partner's explanations can be condescending.
Content Appropriateness (0.72)	I got just the right amount of information from my partner.

3.3.2 Experiment 2 Results & Discussion

According to the first hypothesis, we expected to find that the omission of detail in the Names Only condition would have a negative consequence on information exchange for novices. Results from Experiment 1 support this hypothesis. In Experiment 2, we found the same pattern of results found in Experiment 1 (see Figure 5). When detail was kept from novices, they had to ask significantly more questions than did experts in the same condition; there was no difference between experts and novices in the Names Plus Description condition ($F [3, 47] = 10.9, p < .01$).

Our second hypothesis was that experts and novices would evaluate the robot more positively if the robot's dialogue matched their level of expertise than if it did not. We expected to see significant statistical interactions for the variables related to social relations. On several key social dimensions, this hypothesis was supported by the data in Experiment 2.

Three measures of social relationship produced a significant interaction (see Figure 6). First, the robot's authority was perceived differently depending on level of expertise and dialogue condition ($F [3, 47] = 6.3, p < .05$). Participants who conversed

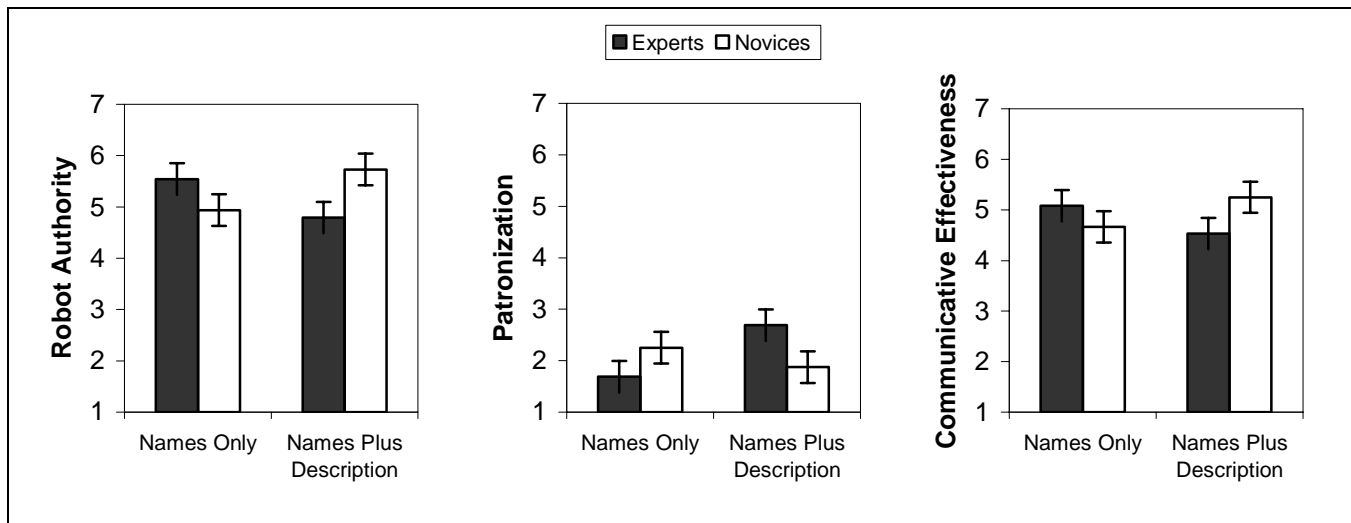


Figure 6. Experts and novices evaluate the robot more positively when the dialogue is adaptive to their information needs.

with the robot whose dialogue matched their level of expertise found the robot to be more authoritative than participants who conversed with a robot whose dialogue did not match their expertise. Thus, experts who interacted in the Names Only condition, and novices who interacted in the Names Plus Description condition thought the robot was more authoritative.

Also, participants conversing with a robot whose dialogue matched their expertise thought the robot was less patronizing than a robot with mismatched dialogue ($F [3, 47] = 4.5, p < .05$). Finally, the questionnaire measure of communicative effectiveness, which included items like “Our conversation was successful,” also showed a significant interaction ($F [3, 47] = 10.97, p < .01$). (See Figure 6.)

Other measures of social relations achieved only marginal significance but were in the expected direction. That is, the appropriately matched robot was marginally seen as more responsive ($F [3, 47] = 3.03, p = .08$) and to have provided more appropriate content ($F [3, 47] = 2.79, p = .10$). In the same manner, those who interacted with the robot whose dialogue matched their expertise also tended to enjoy the task more and to be more willing to participate again ($F [3, 47] = 2.49, p = .12$).

4. GENERAL DISCUSSION

We conducted two experiments testing the effects of an adaptive versus a nonadaptive robot on information exchange and social relations. When the robot used a simple dialogue that pointed out cooking tools using their names, this dialogue was appropriate for experts (who knew about the tools) but not for novices. When the robot elaborated its description of the tools, this dialogue was appropriate for novices but not for experts. We showed that appropriate dialogue improved information exchange for novices and made no difference for experts. Further, when people were under time pressure, the adaptive dialogue improved social relations for both novices and experts. These results suggest that adaptation in human-robot interaction has consequences for both task performance and social cohesion. It also suggests that people may be more sensitive to social relations with robots when communication inefficiencies have actual consequences, as they did in Experiment 2.

4.1 Limitations

Certain tasks for which adaptive dialogue would be advantageous might be better suited to a speech-only interface. While the robot in these experiments does speak aloud, it does not respond to spoken input. We can only speculate that the same effects would be replicated with a speech-only interface.

The robot did not classify experts and novices in these studies. We can only speculate that the same effects would apply if the robot asked the initial test questions, for example. We also did not study other ways of differentiating experts from novices. Moreover, we used a strategy of comparing the extremes of the distributions, leaving out people with moderate cooking expertise. Both of these aspects of the study limit the generalizability of the results and need to be examined further. Ideally, a robot would be able to understand and adapt to many gradations of expertise and user knowledge.

Determining the best way for a robot to appropriately classify individual expertise requires further investigation. For instance,

we used eight questions on cooking terms. The robot could use these and employ a heuristic for determining the split of expert versus novice users. When expertise is displayed for a particular kind of knowledge, the likelihood of knowing other things also changes. In this case, knowing certain cooking terms predicts that the names of cooking tools will also be known. More work needs to be done in other domains of expertise regarding the advantages of stereotyping users for the ease of human-robot interaction.

4.2 Significance

Despite its limitations, we believe this work suggests an important direction for future technical development and social analysis in human-robot interaction. We believe robots that interact with people with diverse needs will be more productive and effective if these robots can assess expertise and adapt. Our study suggests that adaptive robots will not only have instrumental advantages, for information exchange and efficient communication, but that they may have social advantages as well.

4.3 Future Work

Adaptive natural language dialogue will be a challenge on two dimensions. First, we face the challenge of assessing an individual’s level of knowledge and information requirements. The answer will differ across domains. Thus, in guiding people to locations, the robot can learn if they are employees who would be familiar with landmarks or visitors. In giving information to conference attendees, the robot can be aware of the program and the social network, using an individual’s social connections and authorship to understand if they are newcomers to the conference or well established researchers. The domain of giving advice or support to older adults may be quite different. In this case, the robot might need to assess the user’s physical and cognitive capabilities as well as the social context.

Second, we need to know more about people’s responses to robots that adapt to them based on their individual differences. Related research suggests that people respond positively to being mirrored. People appreciate interactive technologies that exhibit similar personality styles [23] or mimic their gestures [2]. We have studied a different kind of adaptation, however. The robot did not imitate the user’s knowledge; instead, it responded to the user’s information needs. In future research we would like to explore the nonverbal dimensions of adaptation as well. For example, if the robot becomes aware that there is a lack of comprehension, it may raise its voice or use exaggerated enunciation. Assessment of this kind carries some risk that the adaptation may be interpreted as a negative evaluation of the person’s competence and could become insulting. Further research should illuminate ways adaptation can be used to maintain social relations between humans and robots while ensuring task success.

5. ACKNOWLEDGMENTS

This research was supported by National Science Foundation ITR project #IIS-0121426 and a National Science Foundation IGERT Fellowship #DGE-0333420 to the first author.

6. REFERENCES

- [1] Atkinson, K. GNU Aspell. <http://aspell.sourceforge.net>.

- [2] Bailenson, J. and Yee, N. Digital Chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological Science*, forthcoming (2005).
- [3] Breazeal, C., Affective Interaction between Humans and Robots. in *European Conference on Artificial Life*, (2001), 582-591, Springer-Verlag.
- [4] Breazeal, C., Brooks, A., Chilongo, D., Gray, J., Hoffman, G., Kidd, C., Lee, H., Lieberman, J. and Lockerd, A. Working Collaboratively with Humanoid Robots. (2004).
- [5] Canary, D.J. and Spitzberg, B.H. Appropriateness and effectiveness perceptions of conflict strategies. *Human Communication Research*, 14 (1987), 93-118.
- [6] Clark, H. and Wilkes-Gibbes, D. Referring as a collaborative process. *Cognition*, 22 (1986), 1-39.
- [7] Clark, H.H. *Using Language*. Cambridge University Press, 1996.
- [8] Fong, T., Nourbakhsh, I. and Dautenhahn, K. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42 (2003), 143-166.
- [9] Fussell, S. and Krauss, R. Coordination of knowledge in communication: Effects of speakers' assumptions about what others know. *Journal of Personality and Social Psychology*, 62, 378-391 (1992).
- [10] Fussell, S. and Krauss, R. Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 21 (1989), 445-454.
- [11] Galinsky, A., Ku, G. and Wang, C. Perspective-Taking and Self-Other Overlap: Fostering Social Bonds and Facilitating Social Coordination. *Group Processes & Intergroup Relations*, 8, 2 (2005), 109-124.
- [12] Giles, H., Coupland, N. and Coupland, J. Accommodation theory: Communication, context, and consequence. in Giles, H., Coupland, J. and Coupland, N. eds. *Contexts of accommodation: developments in applied sociolinguistics*, Cambridge University Press, Cambridge, 1991, 1-68.
- [13] Gockley, R., Bruce, A., Forlizzi, J., Michalowski, M., Mundell, A., Rosenthal, S., Sellner, B., Simmons, R., Snipes, K., Schultz, A.C. and Wang, J., Designing robots for long-term social interaction. in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, (2005), 2199-2204.
- [14] Goffman, E. On face-work: An analysis of ritual elements in social interaction. *Psychiatry*, 19 (1955), 213-231.
- [15] Holtgraves, T. Face management and politeness. in Holtgraves, T. ed. *Language as Social Action: Social Psychology and Language*, Lawrence Erlbaum, Mahwah, NJ, 2002, 37-63.
- [16] Isaacs, E. and Clark, H. References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116 (1987), 26-37.
- [17] Kanda, T., Hirano, T. and Eaton, D. Interactive robots as social partners and peer tutors for children: A field trial. *Human Computer Interaction*, 19 (2004), 61-84.
- [18] Lenzo, K.A. and Black, A.W. Cepstral. <http://www.cepstral.com>.
- [19] Litman, D.J. and Pan, S., Empirically Evaluating an Adaptable Spoken Dialogue System. in *7th International Conference on User Modeling*, (Banff, Canada, 1999), 55-64.
- [20] McCrosky, J.C. Scales for the measurement of ethos. *Speech Monographs*, 33 (1966), 65-72.
- [21] Moore, J., Foster, M.E., Lemon, O. and White, M., Generating Tailored, Comparative Descriptions in Spoken Dialogue. in *17th International Florida Artificial Intelligence Research Society Conference*, (2004), AAAI Press.
- [22] Nass, C. and Brave, S. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, Cambridge, MA, 2005.
- [23] Nass, C. and Lee, K.M. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology Applied*, 7 (2001), 171-181.
- [24] Nourbakhsh, I.R., Bobenage, J., Grange, S., Lutz, R., Meyer, R. and Soto, A. An affective mobile robot educator with a full-time job. *Artificial Intelligence*, 114, 1-2 (1999), 95-124.
- [25] Pineau, J., Montemerlo, M., Pollack, M., Roy, N. and Thrun, S., Towards robotic assistants in nursing homes: challenges and results. in *Workshop on Robot as Partner: An Exploration of Social Robots, IEEE International Conference on Robots and Systems*, (Lausanne, Switzerland, 2002), IEEE.
- [26] Powers, A., Kramer, A., Lim, S., Kuo, J., Lee, S.-I. and Kiesler, S., Eliciting Information from People with a Gendered Humanoid Robot. in *IEEE International Workshop on Robots and Human Interactive Communication (RO-MAN)*, (2005), 158-163.
- [27] Schober, M. and Brennan, S. Processes of interactive spoken discourse: The role of the partner. in Graesser, A., Gernsbacher, M. and Goldman, S. eds. *The Handbook of Discourse Processes*, Lawrence Erlbaum, Mahwah, NJ, 2003, 123-164.
- [28] Schober, M. and Clark, H. Understanding by addressees and overhearers. *Cognitive Psychology*, 21 (1989), 211-232.
- [29] Sidner, C. and Lee, C. Robots as Laboratory Hosts *Interactions*, 2005, 16-24.
- [30] Trafton, J.G., Cassimatis, N.L., Bugajska, M.D., Brock, D.P., Mintz, F.E. and Schultz, A.C. Enabling Effective Human-Robot Interaction Using Perspective-Taking in Robots. *IEEE Transactions on Systems, Man, and Cybernetics--Part A: Systems and Humans*, 35, 4 (2005), 460-470.
- [31] Wallace, R. A.L.I.C.E. *ALICE Artificial Intelligence Foundation*. <http://www.alicebot.org>.
- [32] Warner, R.M. and Sugarman, D.B. Attributions of personality based on physical appearance, speech, and handwriting. *Journal of Personality and Social Psychology* (1986), 792-799.
- [33] Weimann, J.M. Explication and test of a model of communicative competence. *Human Communication Research*, 3 (1977), 195-213.