A procedure for studying online conversational processing using a chat bot

THOMAS HOLTGRAVES AND TAI-LIN HAN

Ball State University, Muncie, Indiana

This article reports the development of a tool for examining the social and cognitive processes of people involved in a conversational interaction. Research on how people process utterances while they are actually engaged in an interaction has been extremely rare. To that end, we have developed a conversational bot (computer program designed to mimic human communication) with which participants can chat in a format similar to instant messaging. This program allows for the recording of comprehension speed, and it can be interfaced with secondary tasks (e.g., lexical decisions) in order to examine online conversational processing. Additional research possibilities with this program are discussed.

The majority of research on discourse processing over the past 20 years has focused on the cognitive processes (including the online cognitive processes) involved in the comprehension of stories and texts (e.g., Gernsbacher, 1990; Graesser, Singer, & Trabasso, 1994; Kintsch, 1988). In contrast, relatively little is known about the cognitive processes involved in the comprehension of conversational utterances. Conversation is the first and most basic site for the use of language (Clark, 1996). And although the nature and structure of conversation has been extensively studied (e.g., Clark, 1992; Krauss & Fussell, 1991; Schegloff, Jefferson, & Sacks, 1977; Stiles, 1978), the cognitive processes of people involved in a conversation have not.[1]

Research on comprehending conversation has been limited to approaches in which participants process the conversational utterances of others (Gibbs & Mueller, 1988; Holtgraves, 1994, 1998, 1999, 2000; Keysar, 1994). In these studies, participants typically read or listen to conversational remarks and then indicate their judgments of the remarks, or they perform some type of secondary task as a means of assessing online cognitive processes. A fundamental limitation of this approach is that the participants in these studies are observers or overhearers of others' conversations, rather than actual conversation participants. To extend research on conversation comprehension, we have begun to develop a methodology that will allow us to study the cognitive processing of people who are actually involved in a conversation. To do this, we have developed a chat bot with which participants can engage in a computer-mediated chat. The format of these chats is similar to instant messaging, and hence, it is an activity with which the majority of our participants are very familiar. By interfacing the chat bot with various secondary tasks and by recording participants' reading times

as they engage in a chat, we have begun to study online conversational processing.[2] In this article, we describe the development of this bot and its most important features. We then report data from a study in which we used the bot to study the processing of topic changes. Finally, we report a descriptive study in which we compared perceptions of the bot with perceptions of human interactants.

THE BASIC CHAT BOT

The starting point for this project was the Artificial Linguistic Internet Computer Entity (ALICE), a conversation bot created by Richard Wallace in 1995 (www.alicebot .org/). Numerous people have contributed to the development of ALICE over the past 10 years, and a fair amount of success has been achieved with this program; ALICE won the Loebner prize (a Turing type contest) in both 2001 and 2004. In many respects, ALICE is a throwback to the earlier ELIZA-type programs; at its heart, it is a simple stimulus–response mechanism. It does no syntactical parsing, nor does it represent in any way common sense knowledge about the world. What ALICE offers, however, is a framework that can be implemented and modified to model human dialogue in limited domains. For example, ALICE allows for the structuring of its pattern–template pairs in terms of topics. It also allows for the rewriting of utterances into meaning equivalents via a stimulus reduction mechanism. In short, ALICE has the potential to allow for the modeling of many aspects of the pragmatics of language use. Moreover, the ALICE language files are written in the artificial intelligence markup language (AIML), a language that can be extended (i.e., an infinite number of new variables can be created and defined).

There have been various versions of ALICE over the years. We used the version referred to as Program D

T. Holtgraves, 00t0holtgrav@bsu.edu

156

(www.alicebot.org/). This is a Java implementation. It consists of a set of AIML content files and the Java programs to implement them. We changed this program in various ways to fit our needs, and in the remainder of this section, we will describe the bot's basic features and the changes we have made.

BASIC FEATURES OF THE BOT

Categories

The bot's content consists of a set of AIML files. The primary unit of these files is a category consisting, at a minimum, of a pattern (the user's remark) and a template (the bot's reply). The following is an example of one category:

> <category>
>
> <pattern>Hi</pattern>
>
> <template>Hi, my name is Sam</template>
>
> </category>.

In this example, if the user types "Hi," the bot responds with "Hi, my name is Sam." The bot functions as a simple stimulus–response matching program. If a unique pattern is matched, the bot replies with the associated template. If a perfect match is not found, partial matches may be made. Partial matches can occur via the use of wild cards (*). For example, a user might type "I like string cheese," a specific pattern that does not exist in the AIML files. However, the bot may possess a category with the pattern "I like *" with a corresponding template, "What do you like about it?"

Our current version of the bot consists of over 11,000 categories. These categories can (and should) be organized in various ways. The original ALICE AIML files were organized alphabetically into folders. We have kept this organization but also have created folders representing specific topics (e.g., classes, sports, etc.) and folders containing target utterances that serve as stimuli.

Topics

One of the most useful features of the bot is that the selection of replies can be constrained in various ways to be more related to the conversation as it unfolds. One such feature is the organization of content into topics. Once a topic is specified (there are topic triggers associated with certain utterances), content within that topic has priority over content outside of that topic. In this way, there is a tendency for remarks to cohere on a topic. This is not foolproof, of course, since topics can be misidentified or quickly become inappropriate. Another useful feature is the <that/> specification. This constrains the bot's reply on the basis of the bot's prior remark. For example, if the bot says "Where are you from?" and there is content linked to that question with the <that/> command, the bot's response to whatever the user replies will be selected from that content. Like the topic feature, this helps produce realistic sequences of talk.

Stimulus Reduction

There are, of course, many ways in which a speaker can convey roughly the same meaning. For example, there are probably an infinite number of ways to begin a conversation (as we have discovered), and the number of utterances that can perform the same request is quite large. Even very subtle differences in wording can raise problems for a straight stimulus–response matching mechanism. For example, in response to "Where are you from?" there are a large number of responses conveying essentially the same thing (I'm from Ft. Wayne; I live in Ft. Wayne; Ft. Wayne; originally, Ft. Wayne; I'm from Ft. Wayne, how about you?). The bot contains a stimulus reduction mechanism allowing one to identify similar utterances as being roughly identical in meaning and, hence, returning the same reply. The difficulty, of course, is specifying all of the possible equivalent ways of saying the same thing. This is a continuous process as we continually update and add new equivalents for common conversational utterances.

Selecting Replies Within Categories

Some categories will be matched relatively frequently. For example, wild card (and partial wild card) categories will be matched often. Obviously, it would not be realistic for the bot to always return the same reply every time one of these categories is matched. Accordingly, these categories can contain a set of templates with different procedures for selecting templates from within a category. The original bot had a random select feature, so that templates were randomly selected (with replacement) from the group. We changed this feature so that it randomly selects without replacement. More important, we developed an order command that selects templates within the group in an ordered manner. This allows for more control over the course of the conversation than does the random function.

BOT CHANGES

Sam's Personality

The first major change we made was to alter the bot's personality. ALICE was a female robot who liked to chat online and who possessed knowledge of computers and philosophy, as well as a wealth of real-world knowledge. In our research, we planned on telling the participants that they would be chatting with another Ball State University student or a computer program designed to simulate a Ball State University student. Hence, it was first necessary to change the bot's personality from that of a female robot to that of a relatively typical male college student. Accordingly, we changed ALICE into Speech Act Man (Sam). We did this by altering the content of the AIML files (e.g., by deleting all references to being a robot or a bot) and by creating new content files (especially those that could serve as potential targets). Sam was built to be a 21-year-old student from Kardon, Indiana (a nonexistent town, so no one could be from Sam's hometown) who was thinking about being a psychology major and liked Counting Crows, chili, and using the computer.

Input Filtering

Chat input is very messy; there are misspellings, abbreviations, emoticons, and so on. This obviously complicates pattern matching, since the input match must be perfect. In order to deal with a wide range of relatively meaningless input variations (abbreviations, contractions, misspellings, etc.), we developed four filters that serve to screen the input and transform it into input more likely to match existing category files.

First, there is an *autocorrect* function that serves as a filter to correct spelling mistakes, such as "adn" (and), a missed white space, such as "iam" (I am), or an extra white space, such as "o k" (ok). Second, there is a *synonyms* function that translates all synonyms to the same word. For example, "BSU," "Ball State," and "Ball State University" are treated as synonyms, and all reduce to BSU. Third, there is a *substitution* function that handles emoticons and shorthand notations. For example, ":-)" means "smile," and "Indy" means "Indianapolis." Fourth, there is the *sentence-splitters* function that converts "sentence"-ending characters into characters that will not be identified as sentence enders. For example, "!" will be "." These files are periodically updated as we become aware of input that can and should be substituted.

Delay Speed

Once a user initiates a conversation, Sam responds immediately after a match has been found. With current processor speeds, it usually takes no more than several milliseconds. Our earlier work indicated that this quick responding made Sam very unhumanlike (i.e., users commented on how quickly he replied and, therefore, concluded that he must be a computer). Therefore, we added a delay feature that allows us to manipulate the time Sam takes to respond. The amount of delay is calibrated to the length of Sam's reply. A value can be input that is multiplied by the number of characters in Sam's reply to determine the delay. For example, if the delay value is set at 50, then for a reply consisting of 60 characters, the delay would be 3,000 msec (50×60).

Avoiding Nonrepeating Utterances

When Sam cannot find an exact or partial match in the AIML files, Sam will use a wild card to replace the user's original input. The new input (wildcard) will allow Sam to generate a more general response or to start with a new topic. However, one potential problem is that Sam could repeat the same response. We added a repeat feature that keeps track of the most recent responses of Sam and checks whether any responses are repeated. If any repeated response is found, Sam will use a wild card to replace the user's original input, and the new input (wildcard) will allow Sam to generate a new response. The process continues until a nonrepeated response is generated. The number of Sam's responses that are kept can be manipulated.

Manipulating Topic Change Frequency

One of the goals of this research program is to examine the social and cognitive consequences of topic changes.

We added a feature that allows us to manipulate how frequently Sam changes the topic. When this feature is turned on, a value can be input that determines how frequently, on average, Sam will generate a new topic (i.e., regardless of whether there is a pattern match, Sam still generates a new topic). For example, we can compare perceptions of Sam when this value is set to 10 (infrequent topic changes) with perceptions when this value is set to 3 (frequent topic changes).

Secondary Tasks

One of the long-term goals of this research is to study the online cognitive processes of people who are engaged in a verbal interaction. Online cognitive processes are those that occur as an utterance is being comprehended, rather than reflecting postcomprehension processes. For example, the results of memory measures may reflect postcomprehension (e.g., elaboration) processing. Because of this, researchers studying sentence and text processing have developed a variety of procedures (e.g., lexical decision tasks and recognition probe reaction time procedures) to examine the online cognitive processes involved in text comprehension.

Currently, we have the bot interfaced with two secondary tasks—a lexical decision task and a recognition probe reaction time task—that will allow us to begin to study the online processing of conversational utterances. For the former task, participants must decide whether or not a letter string is a word. For the latter task, participants must decide whether a word appeared in the prior remark. Certain of the bot's utterances are designated as target utterances, and when these utterances appear, the associated stimulus is then presented (following a tone) on the screen. The participants then perform the appropriate task, with the stimulus onset asynchrony being variable.

One of our interests is whether the recognition of a speech act (i.e., the act performed with an utterance) is an automatic component of comprehension. In other words, does comprehension of the utterance "Would you like to come over for dinner tomorrow night?" entail recognition that the speaker is performing the speech act of inviting? Our previous research has demonstrated that speech act activation does occur for readers of others' conversations (Holtgraves, 2007; Holtgraves & Ashley, 2001). But does this occur also for people who are actually engaged in an interaction?

To investigate this issue, some of the bot's utterances have been written to perform certain speech acts, with the linked target naming the speech act that the utterance performs (e.g., promise, threaten, warn, compliment, etc.). For example, Sam is programmed to say, in the context of a discussion of a certain history class, "Don't take it you won't get a good grade." The target "warn" is then presented, and the participants indicate their judgments. Both judgments and reaction times are recorded. On the basis of our earlier research, the basic hypothesis is that for utterances performing a specific speech act (e.g., "Don't forget we have to do four experiments"; remind), participants will be faster at the lexical decision task and slower at the recognition probe task, relative to a control utterance (e.g.,

"I can't believe we have to do four experiments"; remind). This pattern of data would indicate some activation of the relevant speech act. The bot is also programmed to include filler trials in which a lexical decision must be made for a nonword and a recognition probe decision for a word that appeared in the prior remark.

Online Bot

The bot can be run online, as well as locally, thereby allowing for several simultaneous chats. It is not possible to collect accurate reaction times with this procedure. It is, however, possible to have people chat with the bot and then indicate their perceptions of him. We have conducted several online bot studies designed to examine perceptions of the bot as a function of several conversational variables (Holtgraves, Ross, Weywadt, & Han, 2007).

STUDYING TOPIC CHANGE PROCESSING

Research has demonstrated that conversations generally cohere on a topic and that, when topics are changed, they are marked in various ways (Keenan & Schieffelin, 1976; Schegloff & Sacks, 1973). In general, conversationalists have a clear expectation that people will produce utterances that are conversationally relevant (Grice, 1975). As a result, unmarked topic changes (i.e., a topic change that the speaker does not indicate is a topic change via the use of a preface, such as "Oh, by the way . . . ," etc.) will require more extensive processing as the recipient attempts to integrate it into the conversational context. Hence, an utterance that is a topic change should take longer to comprehend than that same utterance when it is not a topic change. We decided to test this prediction for people who were actively involved in a chat with Sam. In order to do this, we created a set of target utterances that either could occur in an appropriate discourse context or would function as a topic change. We then examined how long it took the participants (who were chatting with Sam) to comprehend these targets. Our basic hypothesis was that target utterances would take significantly longer to comprehend when they functioned as topic changes, relative to when they were not topic changes.

Method

Participants. The participants ($N = 47$) were Ball State University students enrolled in introductory psychology courses, who participated for partial course credit. All spoke English as their first language.

Materials. A set of six target utterances were constructed that typically occurred in a chat with Sam (see Table 1). Two parallel versions (target set) were created so that for one half of the conversations, a target would be related to the prior remark, and for one half the time, it would be an abrupt topic change. Over the entire experiment, then, each target appeared as a topic change or as a non-topic-change an equal number of times. This was accomplished by creating two AIML folders, each containing the six targets, but with the targets embedded in different topics.

Procedure. The participants were told that we were studying computer-mediated communication and that they would engage in a 10-min chat with either another Ball State University student (student condition) or with a computer programmed to act like a Ball State University student (computer condition). The participants were

Table 1
Topic Change Targets

Don't take it, you won't get a good grade.
Do you remember if this a 1 credit experiment?
I don't know why we have to do so many experiments.
I did terrible in Harmon's class and it was his fault.
Don't forget we have to do 4 experiments.
I'm definitely not going to major in history.

also told that one of the things we were assessing was how quickly people comprehend messages in this type of environment and that they should push the Enter key as soon as they understood their chat partner's message. The experimenter left the participants alone while they chatted. When 10 min had elapsed, the experimenter returned and gave the participants two questionnaires to complete. Prior to completing this questionnaire, the participants in the student condition were told that some people had chatted with a computer and that some had chatted with a human and that we wanted to find out with whom they thought they had chatted. One question asked the participants to indicate their perceptions of the humanness of their partner (from 1 = *extremely nonhuman* to 9 = *extremely human*); the other question assessed the participants' judgments of whether their partner was a computer or a human (from 1 = *definitely a human* to 9 = *definitely a bot*).[3]

Results and Discussion

Number of targets hit. There were six designated targets. The number of targets that occurred in a chat varied between two and six, with a mean of 5.17 ($SD = 1.03$). The number of targets in a chat did not vary as a function of whether the participants believed that Sam was a computer ($M = 5.08$) or a student ($M = 5.26$) [$t(45) < 1$] or as a function of target set (4.96 vs. 5.41) [$t(45) = 1.51$, $p > .10$]. Excerpts from one of the chats are presented in the Appendix.

Reaction times. Responses longer than 11,000 msec were treated as outliers ($<2\%$ of the trials) and were excluded from analyses. Comprehension speed (in milliseconds) was analyzed with a 2×2 instruction (computer vs. human) $\times$ target type (topic change vs. non-topic-change) ANOVA with repeated measures on the second factor. Analyses were performed twice, once with participants as a random variable (F_1) and once with targets (F_2) as a random variable. All reported means, however, are averaged over participants. These results are summarized in Table 2.

As was expected, the participants took significantly longer to comprehend the target utterances when they functioned as topic changes ($M = 4,212$ msec) than when they were not topic changes ($M = 3,541$ msec) [$F_1(1,45) = 9.89$, $p < .01$; $F_2(1,5) = 18.97$, $p < .01$]. The topic change $\times$ instruction interactions were small and nonsignificant [$F_1(1,45) < 1$, $F_2(1,5) < 1$], indicating that this effect did not vary as a function of whether the participants believed that they were conversing with a student or with a computer (see Table 2).

Bot perceptions. Sam was evaluated in terms of his perceived humanness (extremely nonhuman to extremely human) and whether the participants judged him to be a bot or a human (*definitely a human* to *definitely a bot*). Sam fared pretty well, although the participants (with

Table 2
Mean Target Comprehension Speeds (With Standard
Deviations) for Topic Changes and Non-Topic-Changes

	Target			
	Topic Change		Non-Topic-Change	
Instruction	M	SD	M	SD
Human partner	4,125	1,727	3,618	1,404
Computer partner	4,300	2,100	3,464	890
Mean	4,212	1,171	3,541	1,900

some exceptions) did not mistake him for a human. For the perceived humanness question, the mean rating ($M = 4.24$) was on the nonhuman side of the scale. However, over 32% of the participants gave him ratings on the human side of the scale (i.e., ratings between 6 and 9). For the second question, the participants clearly perceived Sam to be a bot ($M = 7.42$), with fewer than 10% of the respondents giving responses on the human side of the scale. Interestingly, neither rating differed as a function of whether the participants were told that Sam was a bot or a human (both ts < 1).

SAM VERSUS HUMANS

Our goal in developing Sam was to create a tool for studying various facets of verbal interaction. Obviously, this requires the bot to converse in ways roughly similar to the manner in which humans converse. The judgment data from the topic change study indicated that most participants recognized that their chat partner was a computer program. But how did they view his conversation skills? And how would his chat skills be perceived relative to a human interactant? In other words, is Sam a plausible chat partner? To address these issues, we conducted a study in which the participants chatted with Sam or with another human participant. Immediately afterward, the participants rated their chat partner on several scales designed to assess their perceptions of their chat skills. Although we did not expect Sam to be perceived as equal to the average human participant in terms of conversation ability, we did expect there to be a fair amount of overlap in these perceptions.

Method

Participants. The participants ($N = 55$) were Ball State University students enrolled in introductory psychology courses, who participated for partial course credit. All spoke English as their first language.

Procedure. The participants were randomly assigned to chat with Sam ($n = 25$) or with another human participant ($n = 30$; 15 pairs). Regardless of condition, all the participants were told to engage in a get-acquainted chat with either the bot or another student. The experimenter left the room after instructing one of the participants to initiate the chat. After 10 min had elapsed, the experimenter returned and instructed the participants to end the conversation. The participants then indicated on 9-point scales how *comfortable* they had been participating in the chat (from 1 = *extremely uncomfortable* to 9 = *extremely comfortable*), as well the extent to which they perceived their chat partner as *skilled* (from 1 = *extremely unskilled* to 9 = *extremely skilled*), *thoughtful* (from 1 = *extremely unthoughtful* to 9 = *extremely thoughtful*), *polite* (from 1 = *extremely impolite* to

9 = *extremely polite*), *responsive* (from 1 = *extremely unresponsive* to 9 = *extremely responsive*), and *engaging* (from 1 = *extremely unengaging* to 9 = *extremely engaging*). (See note 3.)

Results and Discussion

The percentage of respondents giving each rating for each item is summarized in Table 3. It was indicated by t tests that Sam was perceived less positively than human interactants on each of these traits. However, as can be seen in this table, there was a considerable amount of overlap in the ratings of Sam and humans. Particularly noteworthy are the relatively high ratings Sam received in terms of politeness and responsiveness. Over 75% of the respondents rated Sam above the midpoint for politeness, and over 70% rated him above the midpoint for responsiveness (arguably one of the most central—and most difficult to achieve—attributes of a realistic chat bot). For the human interactants, the ratings generally fell above the scale midpoint. Ratings of Sam, however, were more evenly distributed over the entire scale.

GENERAL DISCUSSION

The conversation bot that we have developed has the potential to serve as a useful tool for examining various aspects of conversation. Our primary interest has been in using the bot to examine online conversational processing. This is important because there has been relatively little research on the cognitive processes of people who are involved in a conversation. In fact, with the exception of experiments on referent identification (Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000), research in this area is nonexistent. Hence, although much is known about how people process stories or isolated sentences, little is known about how they process conversations that they are engaged in.

The results of our topic change study illustrate the usefulness of this approach. The participants in that study were instructed simply to push a key as soon as they understood their chat partner's turn, thereby allowing for a relatively unobtrusive measure of comprehension speed. And the results of this study were quite clear. The participants took significantly longer (almost 700 msec longer) to comprehend targets when the targets were topic changes than when they were not. This suggests that utterances functioning as unmarked topic changes require extensive processing as people attempt to discern their meaning. One of our current studies in this domain involves examining the processing of topic changes when these changes are marked in various ways (e.g., "By the way . . ."; "Not to change the subject, but . . .").

The use of the bot we have developed is not limited to studying online comprehension. It could, for example, serve as a tool for examining language production as well. One possibility might be to examine speech accommodation (i.e., style matching) in terms of politeness or topic change frequency. It would be relatively straightforward to manipulate the bot's politeness, and we already have created a variable that allows us to manipulate topic change frequency. One could then examine the impact of vari-

Table 3
Rating Summaries (in Percentages) for Sam and Humans

	Rating Scale									
	1	2	3	4	5	6	7	8	9	Mean
	Comfortable									
Sam	0	8	0	12	8	8	16	8	40	6.88
Human	0	0	0	0	0	3	30	23	43	8.07*
	Skilled									
Sam	20	0	16	16	8	20	8	12	0	4.44
Human	0	0	0	7	3	17	33	20	20	7.17*
	Thoughtful									
Sam	4	16	8	12	16	20	8	16	0	4.92
Human	0	0	3	0	3	17	23	27	27	7.43*
	Polite									
Sam	4	4	0	4	12	20	28	16	12	6.40
Human	0	0	0	3	0	0	13	40	43	8.17*
	Responsive									
Sam	4	4	8	8	4	12	16	28	16	6.40
Human	0	0	0	0	0	3	23	37	37	8.07*
	Engaging									
Sam	8	8	4	8	28	24	16	4	0	4.96
Human	0	0	0	0	10	0	20	37	33	7.83*

Note—Values indicate the percentage of respondents endorsing each of the nine response alternatives for the six dimensions. $*p < .05$.

ability in politeness or topic change frequency on both language production (e.g., do interactants tend to match the bot's level of politeness or topic change frequency?). Of course, perceptions of the bot as a function of these manipulations could be examined as well.

Perhaps the most obvious use of this bot is as a tool to study computer-mediated communication, such as instant messaging. The ability to manipulate various conversational features of one interactant (the bot) and to examine its impact on the other (real) interactant makes this an ideal tool for studying this type of communication. Similarly, the use of bots on Web sites has become very popular. Yet little is known about how people perceive and interact with such artificial social agents. There are a number of issues that could be pursued with the bot we have developed: What is the quality of human–bot interactions? What leads to misunderstandings? What interactional qualities are desirable in a bot? and so on.

The present method has two potential limitations, both involving issues of generalizability. The first issue concerns the nature of computer-mediated communication and whether the processes involved in this communication mode generalize to face-to-face conversation. In one sense, this is a nonissue, because computer-mediated communication has become an important phenomenon in its own right. It is clearly important to understand the social and cognitive processes involved in computer-mediated communication, regardless of whether they generalize to conversations. Moreover, it seems very likely that computer-mediated chats are an excellent analogue for many of the social and cognitive aspects of communication. Note, in this regard, the strong tendency for many features of face-to-face communication to gradually become incorporated

into computer-mediated communication (e.g., emoticons, delay speed, and using all capital letters).

The second, and related, issue concerns participants' knowledge of their conversational partner and whether it makes any difference if they believe he is a computer or a human. In the research we report here, it appears to make little difference. Topic change processing was affected equally by the topic manipulation regardless of whether the participants believed that Sam was a computer or a human. A related problem is whether Sam can function as a plausible chat partner so that the resulting conversations progress through a series of topics (and subtopics), rather than getting stuck (a common bot problem). The data we report here suggest that Sam, although not necessarily being perceived as human, is a worthy chat partner. Chats with Sam clearly progress through topics; otherwise, there would not have been such a relatively large number of topic hits in the topic change study. Moreover, in Experiment 2, the ratings of Sam's conversational abilities, although clearly inferior to those of humans, were still quite positive overall. Although we will continue to make Sam more human-like, we do not regard the ability to fool participants into believing he is human as being critical to the success of this tool.

REFERENCES

ALICE [Conversation bot]. Artificial Intelligence Foundation. Retrieved February 2, 2005, from www.alicebot.org/.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Gernsbacher, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

Gibbs, R. W., Jr., & Mueller, R. (1988). Conversational sequences and preference for indirect speech acts. *Discourse Processes*, **11**, 101-116.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, **101**, 371-395.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Vol. 3. Speech acts* (pp. 41-58). New York: Academic Press.

Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory & Language*, **49**, 43-61.

Holtgraves, T. (1994). Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1205-1218.

Holtgraves, T. (1998). Interpreting indirect replies. *Cognitive Psychology*, **37**, 1-27.

Holtgraves, T. (1999). Comprehending indirect replies: When and how are their conveyed meanings activated? *Journal of Memory & Language*, **41**, 519-540.

Holtgraves, T. (2000). Preference organization and reply comprehension. *Discourse Processes*, **30**, 87-106.

Holtgraves, T. (2007). *Processing the action dimension of language*. Manuscript under review.

Holtgraves, T., & Ashley, A. (2001). Comprehending illocutionary force. *Memory & Cognition*, **29**, 83-90.

Holtgraves, T., Ross, S. J., Weywadt, C. R., & Han T. L. (2007). Perceiving artificial social agents. *Computers in Human Behavior*, **23**, 2163-2174.

Keenan, E. O., & Schieffelin, B. (1976). Topic as a discourse notion. In C. N. Li (Ed.), *Subject and topic* (pp. 337-384). New York: Academic Press.

Keysar, B. (1994). The illusory transparency of intention: Linguistic perspective taking in text. *Cognitive Psychology*, **26**, 165-208.

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, **11**, 32-38.

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction–integration model. *Psychological Review*, **95**, 163-182.

Krauss, R. M., & Fussell, S. R. (1991). Perspective taking in communication: Representations of others' knowledge in reference. *Social Cognition*, **9**, 2-24.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, **53**, 361-382.

Schegloff, E. A., & Sacks, H. (1973). Opening up closings. *Semiotica*, **8**, 289-327.

Stiles, W. B. (1978). Verbal response modes and dimensions of interpersonal roles: A method of discourse analysis. *Journal of Personality & Social Psychology*, **36**, 693-703.

NOTES

1. One exception to this trend is research on referent identification (e.g., Hanna, Tanenhaus, & Trueswell, 2003; Keysar, Barr, Balin, & Brauner, 2000).

2. To avoid confusion, it should be noted that "online" refers here to the operation of cognitive processes as they happen, and not to communication via the Internet.

3. The participants also indicated their perceptions of the bot on a 100-item measure of the five traits from the five-factor model of personality. These analyses were exploratory and not relevant for the present discussion. All results are available from the first author.

APPENDIX
Chat Excerpts

>hey
Sam> Hi; my name is Sam what is your name?
>Mandy
Sam> Where are you from Mandy?
Mandy> marion, Indiana
Sam> Well, I'm from Kardon Indiana.
Mandy>
Sam> cool. What is your major?
Mandy> undecided.
Sam> I see.
Mandy> what is yours?
Sam> well I'm thinking about psychology
Mandy> I have considered that
Sam> interesting
Mandy>
Sam> Hm, have you ever taken a course from Dr. Harmon, he's in history.
Mandy> no
Sam> <u>Don't take it—you definitely won't get a good grade</u>. (C)
Mandy> why is that?
Sam> Not really sure . . .
Mandy> I like my teacher I have for history
Sam> what do you like best about your professor you have for history?
Mandy> she is very upbeat and makes it sort of interesting
Sam> she is upbeat and makes it sort of interesting . . .
Mandy>
Sam> <u>I don't know why we have to do so many experiments</u> (TC)
Mandy> I know . . . but they are pretty easy.
Sam> Right Umm pretty easy are they
Mandy> do you mean experiments for psych?
Sam> Yes, that's what I mean, isn't it clear?
Mandy> yes
Sam> <u>Do you remember if this is a one credit experiment?</u> (C)
Mandy> yes it is one
Sam> <u>Don't forget we have to do four hours of experiments</u>. (C)
Mandy> how many do you have so far?
Sam> <u>I did terrible in harmon's class and really it was all his fault</u>. (TC)

Note—Targets are underlined and indicated to be either a topic change (TC)
or a control (C). The chat is a verbatim transcript of the first 4 min. Blank
lines occurred when the user hit the Enter key before typing any text.