

Effects of Message Style on Users' Attributions toward Agents

Susan E. Brennan and Justina O. Ohaeri

Department of Psychology
State University of New York
Stony Brook, NY 11794-2500
Email: brennan@psych.stanford.edu, johaeri@ccvm.sunysb.edu

ABSTRACT

We used a Wizard-of-Oz paradigm to study effects of message style on dialog and on people's mental models of computer agents. People made airline reservations using a simulated reservation agent from which they received one of three message styles: *Telegraphic*, *Fluent*, or *Anthropomorphic*. The agent accepted any kind of language or command input people typed. When people took the initiative, they tended to model their inputs on the computer's messages. They expended more effort in the Anthropomorphic than in the Fluent or Telegraphic conditions. We found no evidence that natural language messages caused higher expectations of intelligence than telegraphic messages.

KEYWORDS: Natural language interfaces, error messages, agents, anthropomorphism, mental models

INTRODUCTION

Do natural language (NL) interfaces lead people to mistake computers for social beings or for overly intelligent ones? This charge has been levied against NL interfaces by several critics (Schneiderman, 1992, 1993; Friedman and Kahn, 1992). Perhaps NL interfaces have not realized their full potential in the interface because they cause users to form unrealistic expectations about systems. While anthropomorphic representations have their uses (Laurel, 1991), there could be undesirable consequences if they sometimes lead to errors or deception. Are NL interfaces inherently anthropomorphic? If so, should NL researchers worry?

Another possibility is that NL dialogs can be unpredictable and problematic simply because they offer more degrees of freedom than command or direct manipulation dialogs. Discourse processing and repair have been studied less than other areas of psycholinguistics; if we better understood the forces influencing communication, we might be better able to build NL interfaces that work well in particular domains and that could be integrated with today's direct manipulation interfaces. Research in human communication suggests that people design their utterances with addressees in mind, that conversants

come to use a limited and idiosyncratic vocabulary with one another, and that they even entrain on the style and syntax of one another's utterances. Perhaps these and other dialog behaviors could be exploited to improve NL interfaces.

Our purpose was to see if the style of messages provided by a computer agent during a task-oriented dialog affected users' attributions toward the agent, as made manifest by users' expectations during the dialog and their evaluations afterward. We tested several hypotheses about text-based interfaces: 1) people will model the kind of language used by a computer agent; 2) systems that present complete-sentence responses and error messages will take more effort to deal with than those that present more concise messages; 3) people will (mis)attribute more intelligence to systems that use complete sentence responses; and 4) people will (mis)attribute more intelligence to NL systems that present themselves as anthropomorphic by using first person pronouns in messages.

METHOD

Subjects were 33 students at the State University of New York at Stony Brook, all native speakers of English. They were given six travel scenarios and were told to use the computer to make airline reservations. Unknown to the subjects, the computer partner in this task was a human operator who provided rule-based textual responses from a terminal in another room. Each person was randomly assigned to one of three message conditions: 1) Anthropomorphic, 2) Fluent, and 3) Telegraphic (see Table 1). The first two types of messages were complete grammatical sentences, while Telegraphic ones were not. Anthropomorphic messages used first person pronouns, "I" and "me," whenever possible; the other two did not use these pronouns.

Greeting

Anthro: Welcome to CompuTravel. How may I help you?

Fluent: Welcome to CompuTravel. Please enter your first request.

Telegr: Welcome to CompuTravel. Enter first request.

Response to a typo, misspelling, or unknown word

Anthro: I don't know the word "travdl"

Fluent: The word "travdl" is unknown

Telegr: Unknown word: "travdl"

Prompt for information

Anthro: Tell me where you are departing from.
Fluent: Where are you departing from?
Telegr: Needed: Point of departure.

Table 1. Examples of message styles

In all three conditions the computer "understood" everything the user typed to the extent that a human partner would, except that it did not accept typos and misspellings.

People completed pre- and post-session questionnaires to measure attributions toward the computer. The post-test included judgments of acceptability (on a scale of 1-7) that people made about a list of hypothetical inputs to the computer. We analyzed questionnaires and behavioral data coded from the session transcripts.

RESULTS AND DISCUSSION

People were equally successful in all three groups; out of 6 possible scenarios, they completed an average of 5.4, 5.1 and 4.5 ($F(2,30)=1.63$, n.s.) respectively in the Anthropomorphic, Fluent, and Telegraphic groups. The Anthropomorphic group was more than twice as likely to refer to the computer using the second-person pronoun "you" (6.6 times on average) as the Fluent (2.5) or the Telegraphic (2.7) groups ($F(2,30)=3.98$, $p<.03$). This indicates that the manipulation worked; the Anthropomorphic group appeared to treat the computer more like a social partner than did the others.

With respect to word counts, the Fluent condition resembled the Telegraphic condition more than it did the Anthropomorphic condition. People typed on average 72.3 words per scenario in the Anthropomorphic, 59.8 in the Fluent, and 50.6 in the Telegraphic group ($F(2,30)=3.12$, $p<.06$). Minutes to complete the task showed a similar pattern (65.4 to 52.3, $F(2,30)=2.67$, $p<.09$), though this difference was marginal. The number of words people typed during a session was unrelated to their self-reported typing ability ($r=.148$, n.s.). Anthropomorphic messages led to more indirect requests and politeness conventions (e.g. "please," "thank you") than did Fluent or Telegraphic messages, 21.6 to 13.0 to 7.1 ($F(2,30) = 4.45$, $p=.02$).

As for attributions of intelligence, there was no difference in ratings between the three groups. In no group did people appear to believe that the computer had general knowledge outside of the immediate task domain; in the post-test, they judged off-task inputs (e.g. "Could you suggest some good hotels in the Miami area?") as equally unacceptable to the computer in all three groups. Converging evidence comes from the fact that there was no difference in the number of off-task requests the three groups made to the computer.

There appeared to be two distinct strategies for communicating with the reservation agent: one involved

the user's taking the initiative and specifying each reservation as much as possible, and the other involved letting the agent take the initiative and ask the user for information. Some people seemed to prefer taking the initiative; others preferred letting the computer take control; others alternated strategies. In turns where people took the initiative, Anthropomorphic messages led to more complete-sentence inputs (vs. telegraphic) than did the Telegraphic condition, 32.0 to 15.01, with Fluent falling in between ($F(2,30)=4.45$, $p<.04$). This replicates previous findings that the language people use with computer systems is shaped by the systems' language (Brennan, 1991; Brennan & Lee, 1994; Zoltan-Ford, 1991).

CONCLUSION

To sum up, we found no evidence that people think interfaces that respond in fluent or anthropomorphic-style messages are any more intelligent than interfaces that respond in a telegraphic style. People tended to use indirect phrasing and politeness conventions more often with the anthropomorphic version than with the fluent and telegraphic ones. Since many NL systems cannot interpret indirect queries, the fluent response style may be preferable to the anthropomorphic one if it shapes users to provide input the system can handle.

Dialog initiative may have greater effects on user performance and satisfaction than whether or not an interface uses fluent NL. In fact, one of the reasons that people have had difficulty using NL interfaces in some situations may have been that they were required to take the initiative in a task with which they were inexperienced. In our future studies of NL interfaces, we plan to control for initiative and also to examine whether users are more likely to spontaneously take initiative with an interface that provides fluent NL input and output than with a command interface that provides telegraphic output.

Finally, we believe that it is counterproductive to pit direct manipulation interfaces against natural language interfaces. Each modality has its advantages; we seek to understand how to make the most of both.

ACKNOWLEDGEMENTS

Support was provided by NSF Grant #92-02458, Dept. of Ed. Grant #P202A30062 (Minority and Women Research Apprenticeship Program), and Apple Computer, Inc.

REFERENCES

- S. E. Brennan. Conversation with and through computers. *User Modeling and User-Adapted Interaction*, 1: 67-86, 1991.
- S. E. Brennan and G. Lee. Solving the vocabulary problem in language interfaces. Manuscript, 1994.
- B. Friedman and P. H. Kahn. Human agency and responsible computing: Implications for computer system design. *J. Systems Software*, 17: 7-14, 1992.

- B. K. Laurel (ed.), *The Art of Human-Computer Interface Design*. Addison-Wesley, Reading, MA, 1990.
- B. Schneiderman . *Designing the User Interface* (second edition). Addison-Wesley, Reading, MA, 1992.
- B. Schneiderman. *Beyond Intelligent Machines: Just Do It!* IEEE Software, pp. 100-103, January, 1993.
- E. Zoltan-Ford. *How to get people to say and type what computers can understand*. *Int. J. Man-Machine Studies*, 34:527-547, 1991.