Computers in Human Behavior 49 (2015) 245–250

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations

Jennifer Hill [a,*], W. Randolph Ford [b,1], Ingrid G. Farreras [c,2]

[a] School of Engineering and Applied Science, The George Washington University, 800 22nd St. NW, Washington, D.C. 20052, USA
[b] Data Analytics Program, Graduate School, 3501 University Boulevard East, University of Maryland University College, Adelphi, MD 20783, USA
[c] Psychology Dept., Hood College, 401 Rosemont Ave., Frederick, MD 21701, USA

ARTICLE INFO

Article history:
Available online 20 March 2015

Keywords:
CMC
Instant messaging
IM
Chatbot
Cleverbot

ABSTRACT

This study analyzed how communication changes when people communicate with an intelligent agent as opposed to with another human. We compared 100 instant messaging conversations to 100 exchanges with the popular chatbot Cleverbot along seven dimensions: words per message, words per conversation, messages per conversation, word uniqueness, and use of profanity, shorthand, and emoticons. A MANOVA indicated that people communicated with the chatbot for longer durations (but with shorter messages) than they did with another human. Additionally, human–chatbot communication lacked much of the richness of vocabulary found in conversations among people, and exhibited greater profanity. These results suggest that while human language skills transfer easily to human–chatbot communication, there are notable differences in the content and quality of such conversations.

1. Introduction

Artificial intelligence's (A.I.) efforts in the last half century to model human language use by computers have not been wildly successful. While the idea of using human language to communicate with computers holds merit, A.I. scientists have, for decades, underestimated the complexity of human language, in both comprehension and generation. The obstacle for computers is not just understanding the meanings of words, but understanding the endless variability of expression in how those words are collocated in language use to communicate meaning.

Nonetheless, decades later, we can find an abundance of natural language interaction with intelligent agents on the internet, from airline reservation systems to merchandise catalogs, suggesting that humans have little or no difficulty transferring their language skills to such applications. Because so much of this communication occurs through digital technology rather than in person, computer-mediated communication (or "CMC") has become a prominent area of research in which to explore this simulation of natural human language.

One of the most popular forms of CMC today, particularly among adolescents and teenagers, is instant messaging (IM) (Tagliamonte & Denis, 2008). While many specialized applications enable instant messaging, the service is also provided through many other popular media, such as multiplayer online games, email clients, and social networking websites (Varnhagen et al., 2009).

Several studies have compared IM and other forms of CMC to other forms of language. Ferrara, Brunner, and Whittemore (1991) determined that CMC possesses uniquely distinguishing linguistic features that display qualities of both written and spoken dialogue. Compared to other standard forms of communication, CMC's most distinctive trait is its unique, shortened-form language of acronyms and abbreviations, and an informal discursive style that is similar to face-to-face spoken language (Werry, 1996). CMC differs from spoken communication, however, in its lack of cues from features such as body language, communicative pauses, and vocal tones (Hentschel, 1998). Despite this absence of cues, however, CMC has been found to be able to communicate emotion as well as or better than face-to-face communication (Derks, Fischer, & Bos, 2008).

Although CMC has been compared to other forms of communication, few studies have compared different forms of CMC to one another. Perhaps the most noteworthy of these studies is Baron's (Baron, 2007) comparison of the linguistic characteristics of IM and text (or SMS) messages – another form of CMC – among

* Corresponding author. Tel.: +1 301 693 4579.
E-mail addresses: jenhill@gwu.edu (J. Hill), rand.ford@umuc.edu (W. Randolph Ford), farreras@hood.edu (I.G. Farreras).
[1] Tel.: +1 240 684 5606.
[2] Tel.: +1 301 696 3762.

http://dx.doi.org/10.1016/j.chb.2015.02.026
0747-5632/

American college students, which found that the average text message contained more words, characters, sentences, abbreviations, and contractions than the average instant message.

To our knowledge, however, no research has investigated the linguistic characteristics of a different form of CMC: chatbot communication. Chatbots, or chatterbots, are another widespread domain of CMC. Chatbots are "machine conversation system[s] [that] interact with human users via natural conversational language" (Shawar & Atwell, 2005, p. 489). Users interact with these applications primarily to engage in small talk. Functionally, their approach to natural language processing is an extension of the same technique used in Weizenbaum's ELIZA (Weizenbaum, 1966). A variety of new chatbot architectures and technologies (e.g., Ultra Hal, ALICE, Jabberwacky, Cleverbot) have arisen recently, each attempting to simulate natural human language more accurately and thoroughly (Carpenter, n.d. a; Shawar & Atwell, 2007; Wallace, Tomabechi, & Aimless, 2003; Zabaware., n.d.).

Despite the popularity of chatbots today, we are not aware of any research analyzing how humans converse with them, particularly from a linguistic perspective. Several extant studies on chatbots have focused on developing or improving their ability to interpret and respond meaningfully to human language: one study examined a chatbot's ability to respond correctly when faced with common CMC features like abbreviations and overlapping utterances from multiple speakers (Shawar & Atwell, 2005), while another examined a chatbot's robustness when faced with unconventional linguistic features from non-native ESL speakers (such as misspellings and incorrect word order) (Coniam, 2008). Another area of research has focused on evaluating users' attribution of human qualities or personality traits to the chatbots they converse with, and how that may lead to greater disclosure in medical, research, or therapeutic settings (Hasler, Tuchman, & Friedman, 2013; Holtgraves, Ross, Weywadt, & Han, 2007; Lortie & Guitton, 2011). Lortie and Guitton (2011) specifically investigated how judges go about distinguishing between humans and computers when interacting in a formal Turing Test. They tracked several descriptive and cognitive parameters along with indicators of interest using the Linguistic Inquiry and Word Count (LIWC; Pennebaker, Chung, Ireland, Gonzales, and Booth (2007)) program. Their results suggested that judges' determinations of humanness were associated with communication that contains more words per message, a higher percentage of articles, and a higher percentage of words that were longer than six letters. Such communication, however, is biased by the goal-directed behavior of judges trying to figure out if they are talking to a computer.

2. Current study

The purpose of this study was to investigate how users' explicit and implicit transfer and expectations of human language is manifested in human–computer interaction. Specifically, we sought to answer an unexplored question in the fields of both computer-mediated communication and chatbot development: do humans communicate differently when they know their conversational partner is a computer as opposed to another human being?

To accomplish this, we compared 100 random human IM conversations against 100 random human conversations with Cleverbot. First launched on the internet by Rollo Carpenter in 1997, Cleverbot is one of the most popular chatbots in existence today. It is designed to simulate natural conversation by learning from human interaction, relying on feedback to communicate (Carpenter, n.d. b). Using a synchronous one-to-one online interface reminiscent of an instant messenger, a human user types a statement and then waits for Cleverbot to send a response.

User: Hi.

Cleverbot: Do you like cheese?

User: No.

Cleverbot: Why not?

User: Not a big fan.

Cleverbot: You like animals?

User: Yes. I like cats.

Cleverbot: Do you like me the best though?

User: I have to get to know you more to see.

Cleverbot: Do you like herman cain?

User: No.

Cleverbot: Why?

User: Cause's I dont know nothing about him.

Cleverbot: My name is Cleverbot.

Fig. 1. A typical conversation with Cleverbot.

Cleverbot is unique in that it "learns" what humans have said to it in previous conversations and uses this knowledge to determine how to respond in new conversations. Figure 1 demonstrates a typical conversation with Cleverbot.

Cleverbot passed the 2011 Turing Test at the Techniche Techno-Management Festival held by the Indian Institute of Technology Guwahati. Of the 1334 volunteers who participated in four-minute typed conversations with either Cleverbot or real humans, 59% rated Cleverbot as human while 63% rated the real humans as human (Aron, 2011), suggesting that Cleverbot is one of the most advanced and human-like conversational agents currently available to the public.

In our comparison of 100 random human IM conversations against 100 random human conversations with Cleverbot, we examined the amount of written content, the uniqueness of the words used, the frequency of profane language, and the use of standard CMC linguistic features such as shorthand phrases (e.g., "lol") and emoticons (e.g., ":-)"). We expected the linguistic profiles of the human–human IM conversations to be consistent with those of previous studies. We also hypothesized that people would send fewer messages, write fewer words per message, and exhibit a more limited vocabulary when communicating with a chatbot compared to another person. We expected this based on three characteristics found in human–chatbot conversations that are not found in human–human conversations. First, people have less experience communicating with chatbots than with humans, and so would be less confident and comfortable with the communicative ability of a chatbot. Second, while chatbots are designed to sustain conversations, they are limited in their ability to have an extended goal-directed discussion and in their depth of information on a particular topic. Third, there is no element of common history or shared experience possible in human–chatbot conversations. Even if particular users have had extensive prior conversations with a chatbot, the chatbot is unaware of such interactions and therefore unable to reference them. As a result, we expected human–chatbot conversations to be more tentative, resulting in shorter conversations, shorter messages, and a limited vocabulary as measured by type/token ratios. Furthermore, we expected the anonymous nature of human-machine conversation to lead to a greater use of profanity in human–chatbot compared to human–human conversations.

3. Method

3.1. Participants and procedure

After obtaining approval to conduct this research from the Institutional Review Board, we collected data from two different communication sources. For the human–human conversations we collected IM conversations from a convenience sample of undergraduate and graduate students at a liberal arts college in the mid-Atlantic region. We asked participants to submit unaltered conversations from past IM sessions. All participant names and screen names were replaced with unique user numbers, ensuring the anonymity of the contributors. We only accepted conversations that took place between two people.

For the human–chatbot conversations, Rollo Carpenter, the creator of Cleverbot, provided us with random dialogues from Cleverbot. Users who converse with Cleverbot do so anonymously – as the website does not assign monikers to its users – and are fully aware that they are speaking to a chatbot. We obtained past conversational histories for both the human–human and the human–Cleverbot conversations so as to circumvent the effects of the Observer's Paradox (Labov, 1972).

Volunteers provided a total of 894 IM conversations, taking place between 205 unique users; Rollo Carpenter provided us with 2140 Cleverbot conversations. To maintain a normal distribution of data and to ensure that each conversation contained enough messages to be analyzed effectively, we removed all conversations with fewer than 10 messages. Additionally, we only used conversations held in English.

We randomly selected 100 conversations of each type of communication for analysis. Because we only analyzed the human half of the human–Cleverbot conversations, we stripped the responding party from the human–human IM sessions from each log prior to analysis. Collectively, the two types of conversations contained 7261 messages, with a total of 41,037 words.

3.2. Linguistic variables analyzed

We investigated seven dependent variables in this study. Three of the variables – words per conversation, messages per conversation, and average number of words per message – pertained to the amount of written content within each conversation. The fourth dependent variable was a type/token ratio that analyzed the percentage of unique words per conversation. The remaining three variables compared common IM shorthand (such as abbreviations and acronyms), emoticons, and occurrences of profanity between the two types of communication. The following paragraphs detail how we operationally defined these variables.

3.2.1. Words

A word was considered to be any cluster of letters separated on either side by a boundary of white space. URLs and gibberish strings used as a representation of confusion or frustration (e.g., "kjdhflkuher") were treated as single words. Emoticons and punctuation chains were not included in the word count.

3.2.2. Messages

A message was defined as any single transmission of data from the user to his or her conversational partner, regardless of the length or content of the transmission. A transmission containing multiple sentences and a transmission made up of a single character or emoticon were each considered to be one message. Repeated transmissions of the same word or sentence were each counted on an individual basis.

3.2.3. Type-token ratio

A type-token ratio is an indication of word diversity within each conversation. This number is a percentage that represents the ratio of unique words (or "types") to the total number of words ("tokens") in a given conversation.

3.2.4. Profanity

A list of profane words was compiled from www.banbuilder.com, a web resource designed to generate custom-banned word lists by desired level of intensity (Gianotto, n.d.). At the time of this study, the database contained 814 words, each classified according to type (e.g., swear, sexual, racial) and perceived intensity rating (G, PG, PG-13, or R). We included all profane words with a rating greater than G, creating a lexicon of 610 profane terms.

3.2.5. Shorthand

CMC possesses linguistic and stylistic differences from both speech and written language (Yates, 1996). One of the most notable differences is its specific and informal shortened-form language, including abbreviations and the omission of auxiliary verbs and pronouns, which has developed over time due to the need to express oneself "as quickly and efficiently as possible" (Werry, 1996). However, there is still a fair amount of inconsistency in the more specific classifications of the various forms of CMC-specific language (Craig, 2003; Driscoll, 2002; Varnhagen et al., 2009). Due to the dispute over the appropriate categorization of IM language, we grouped into a single "shorthand" category all words and phrases encompassed by any of these three authors' different classifications.

As there does not appear to be a single reputable resource for compiling such terms, we adapted our list of shorthand terms from www.Internetslang.com, a frequently-updated database of slang terms, acronyms, and abbreviations that claims to be a "complete dictionary of slang" ("Internet Slang Words. n.d.). The database contained 8028 shorthand words and phrases at the time of analysis, including many emoticons. We excluded any terms that are commonly seen in offline written works (e.g., "BOGO" for "Buy One, Get One", or the phrase "knock it off" meaning "stop it"), as well as acronyms used more frequently as actual words than as shorthand for a particular phrase (e.g., the acronym "SLAP" for "sounds like a plan").

3.2.6. Emoticons

"Emotional icons," commonly referred to as "emoticons," are combinations of ASCII characters used to represent a specific emotional state such as happiness or sadness (Baron, 2003). As with shorthand, there is no one reputable lexicon of emoticons, due in part to the ever-changing nature of the internet and its users, but we adapted the list of emoticons used for this study from www.Internetslang.com ("Internet Slang Words n.d.). We identified and extracted by hand emoticons from the database and compiled them into a separate list for tabulation. The list contained a total of 117 emoticons, including the more commonly recognized "horizontal" faces (e.g.,;) and:-o), Japanese-style "vertical" faces known as *kaomoji* (literally "face marks", e.g., ^_^ and >//<) (Katsuno & Yano, 2002), and miscellaneous symbols used to express emotion (e.g., <3 and \O/).

3.3. Variables from the LIWC

In order to facilitate comparisons with the results of Lortie and Guitton (2011), we also investigated 14 variables that were collected using the LIWC program: word count, words per sentence, percentages of words greater than six letters, LIWC dictionary words, function words, pronouns, articles, swear words, social words, affective words, positive emotion words, negative emotion words, sexual words, and words of assent.

3.4. Statistical analyses

We analyzed the seven variables coded from the two corpora using a multivariate analysis of variance (MANOVA) (with 198 degrees of freedom for each dependent measure), and considering $p < 0.05$ as statistically significant. We also analyzed the results of the 14 variables collected using the LIWC program for the two corpora using a multivariate analysis of variance (MANOVA) (with 198 degrees of freedom for each dependent measure).

4. Results

4.1. Linguistic variables

The following are the results of the MANOVA for the seven dependent variables. The Table 1 presents the means, standard deviations, and subsequent univariate ANOVA results analyzing the seven variables by communication type. Of the seven variables, four were statistically significant: messages per conversation, words per message, type/token ratio, and frequency of profanity. There was a significantly greater number of human–chatbot messages and profanity in these messages, but with significantly fewer words per message and a significantly smaller type-token ratio compared to human–human messages.

Messages to chatbots contained fewer words per message (*Range* = 2–13) than those sent to another person (*Range* = 2–25), but people sent more than twice as many messages to chatbots (*Range* = 19–248) than to other people (*Range* = 3–122). Half of all human–human conversations were 15 or fewer lines long, while none of the human–chatbot conversations were that short.

To test whether people were simply modeling their communication to match that of the chatbot's, we went back and analyzed both the words per message and the type/token ratio for the *responding* half of each human–chatbot and human–human conversation that we had stripped from the original analysis. Humans conversing with Cleverbot expressed a mean 4.29 words per message, close to Cleverbot's mean 3.57 words per message. In contrast, humans conversing with another human averaged 7.95 words per message, while their responding human partners averaged 8.14 words per message. Furthermore, the mean type-token ratio of 47% which we had originally found for humans conversing with Cleverbot almost matched Cleverbot's ratio of 49%, while the 67% type-token ratio we found for initiating humans closely resembled their responding human partners' ratio of 63%.

There was also an almost 30-fold greater average use of profanity in the human–chatbot conversations compared to human–human conversations (4.29% of the conversation vs. 0.16%; *Range* = 0–20% vs. 0–2% of the conversations, respectively). In 85% of the human–human conversations, no foul language appears at all, while only 20% of human–Cleverbot conversations contained no foul language.

The remaining three variables were not statistically significant: number of words per conversation, shorthand terms, and emoticons. There were no statistically significant differences in the mean number of words per conversation, mean number of shorthand terms per conversation, and mean number of emoticons used per session in the human–human vs. human–chatbot conversations.

4.2. LIWC variables

The MANOVA that examined the difference among 14 variables analyzed by the LIWC program between both communication types found a statistically significant effect: *Lambda* (14,185) = 23.076, $p = 0.001$, $\eta^2 = 0.636$. Table 2 presents the means, standard deviations, and follow-up univariate ANOVA results on the 14 variables.

The subsequent univariate ANOVAs on the 14 variables identified eight variables that significantly discriminated between both communication types. People used significantly more pronouns, more swear words, more social words, more negative emotion words, and more sexual words when communicating with Cleverbot as opposed to another person. They used over nine times more words per message, longer words, and more positive emotion words when communicating with another person as opposed to Cleverbot, however.

5. Discussion

The purpose of this study was to compare human to human conversations with human to chatbot conversations. One hypothesis was that the average human would send fewer messages and write fewer words per message when sending messages to chatbots than when communicating with other humans. While messages sent to chatbots did contain fewer words per message than those sent to another person, as predicted, people were actually inclined to send more than twice as many messages to chatbots compared to other people, contrary to our expectations and disconfirming the notion that people feel less confident or comfortable communicating with chatbots. The most reasonable explanation for this finding – combined with the unexpected finding of greater messages to chatbots – seems to be not that people were tentative in their communication with chatbots, but rather that they were modeling their communication to match that of the chatbot's, in the same way that people adapt their language when conversing with children (Bloom, Rocissano, & Hood, 1976; Hausendorf, 1992) or non-native speakers (Ferguson, 1975).

To test this hypothesis, we went back and analyzed both the words per message and the type/token ratio for the *responding* half of each human–chatbot and human–human conversation that we

Table 1
The means, standard deviations, and univariate ANOVA results of the seven dependent variables measured in human–human and human–chatbot conversations.

	Human (N = 100) M (SD)	Chatbot (N = 100) M (SD)	F	p	η^2
Words/conversation	190.42 (212.61)	219.95 (201.95)	1.01	0.315	–
Messages/conversation	23.03 (22.33)	49.58 (37.11)	37.58	0.001*	0.16
Words/message	7.95 (4.02)	4.29 (1.64)	70.86	0.001*	0.26
Type-token	0.67 (0.14)	0.47 (0.09)	144.40	0.001*	0.10
% Profanity	0.16 (0.39)	4.29 (4.55)	81.65	0.001*	0.29
Shorthand	2.69 (4.17)	2.28 (4.03)	0.50	0.480	–
Emoticons	0.70 (1.37)	1.01 (2.90)	0.93	0.335	–

* Statistically significant.

Table 2
The means, standard deviations, F values, degrees of freedom, p values, and effect sizes for the 14 variables analyzed by the LIWC program between human–human and human–chatbot conversations.

	Human (N = 100) M (SD)	Chatbot (N = 100) M (SD)	F	P	η^2
WC	191.12 (212.00)	219.43 (202.42)	0.93	0.335	
WPS	42.77 (114.85)	4.65 (2.03)	11.00	0.001*	0.05
Six-letter	10.50 (4.96)	8.25 (3.38)	13.97	0.001*	0.06
Dic	87.49 (5.51)	86.72 (6.77)	0.77	0.380	
Funct	53.77 (7.58)	54.75 (8.29)	0.76	0.384	
Pronoun	17.25 (4.29)	22.90 (4.99)	73.69	0.001*	0.27
Article	4.50 (2.37)	4.21 (2.31)	0.76	0.382	
Swear	0.23 (0.67)	1.60 (2.14)	37.08	0.001*	0.15
Social	10.61 (5.75)	16.40 (4.73)	60.45	0.001*	0.23
Affect	9.68 (6.01)	8.95 (4.55)	0.93	0.335	
Pos Emot	7.78 (5.64)	6.00 (3.92)	6.67	0.011*	0.03
Neg Emot	1.89 (1.99)	2.94 (2.56)	10.41	0.001*	0.05
Sexual	0.28 (1.03)	3.45 (4.25)	52.51	0.001*	0.21
Assent	3.84 (3.32)	3.65 (3.01)	0.17	0.676	

* Statistically significant.

had stripped from the original analysis. Humans conversing with Cleverbot expressed a similar (small) number of words per message as Cleverbot did, while humans conversing with another human averaged a similar (higher) number of words per message as their responding human partners. The mean type-token ratio of humans conversing with Cleverbot was also quite similar to Cleverbot's ratio, while the type-token ratio we found for initiating humans closely resembled that of their responding human partners' ratio. Both findings seem to confirm this ad-hoc adaptation hypothesis, paving the way for a new and fertile area of research wherein chatbots are used to determine people's ability to alter communication so as to match that of another conversant.

Another hypothesis of this study was that the anonymous nature of human-machine conversations would lead to a greater use of profanity when sending messages to chatbots than when communicating with other humans. The almost 30-fold average increase in profanity in the human–chatbot conversations compared to human–human conversations confirmed our hypothesis and is consistent with prior research (Jay, 2009), although the disproportionate use of profanity in the human–chatbot conversations exceeded our expectations. Beyond mere profanity, many of the conversations were sexually explicit. Were users testing the limits of the chatbot's conversational domains? We have no way of knowing or interpreting users' motivations for such interactions.

We included an analysis of selected dependent variables generated by the LIWC program to draw some parallels between the results of our study and the study by Lortie and Guitton (2011). In their work, they attempted to uncover the features of language that judges may have been using to make their Turing Test judgments. Such conclusions, as the authors were obviously aware, must be drawn cautiously since the very language that was being analyzed may have been confounded by the judgment process. Specifically, human users, either consciously or unconsciously, may have altered their language because they knew they were being evaluated for "humanness." Moreover, the communication of judges in that study may have also influenced the results of the LIWC analysis, as judges were in the process of making a Turing Test decision. The results of the current study afforded us an opportunity to evaluate the validity of the features that may have led to decisions by the Turing Test judges.

In fact, there is a noteworthy degree of agreement between the apparent perceptions of Turing Test judges, and the features that distinguish human–human from human–chatbot communication. Based on the LIWC analyses, people employ more words, longer words, and positive emotion words when communicating with other people than with a computer. People communicating with Cleverbot used higher percentages of swear words, negative emotion words, and sexual words, corroborating our preliminary analysis of profanity use. We did not find, however, the significantly higher percentage use of articles in human–human communication that judges in the Lortie and Guitton (2011) study suggested indicated humanness.

6. Limitations

There were several limitations to this study. Because our human–human corpus was derived from conversations submitted by volunteers, it is difficult to assess how representative they were of natural human conversation. Although we ensured complete anonymity to all volunteers, there is no guarantee that the conversations submitted were not first censored or otherwise altered in some way. Additionally, the human volunteers were not restricted to submitting a single conversation, so our corpus of IM logs could contain several conversations from the same volunteer, whereas each Cleverbot conversation is assumed to have involved a unique speaker due to its massive user base.

Because the conversations were submitted anonymously by student volunteers from a single college, no demographic information was obtained, nor could that information be compared against the more international demographic characteristics of Cleverbot users.

In addition, the LIWC analysis s was limited in that it did not include an analysis of language generated solely by Cleverbot for the purpose of comparison with the Lortie and Guitton (2011) article.

7. Future research

Follow-up research could include a parts of speech analysis that would allow for a detailed comparison of the grammatical structures of the conversations in each corpus. This analysis should isolate individual sentences in both seriatim fragmented messages and messages containing multiple sentences. Another area of research could investigate the differences in conversation length by analyzing the amount of physical time spent in each conversation as well as the length of silences between individual messages. A third area of research could evaluate the number of times an individual conversation shifts topics, and compare this to the total number of messages within that conversation.

8. Conclusion

Whenever we interact with our environment or with other people, we put to use our adaptive processes. This is how we can walk in our house in the dark, search for items in an unfamiliar grocery store, or converse with a stranger. While these adaptive processes allow us to successfully negotiate the novel and unexpected events in our lives, they nonetheless come with a cost. Such situations require that we pay more attention, draw more from the history of our experiences, and be ready to change tactics quickly. In such situations, the burden of this overhead of effort may eventually exceed the perceived value of the activity itself.

When we originally formed our hypotheses for this study, we were mindful of this overhead. It seemed reasonable that the effort required to communicate with a chatbot, an entity both unfamiliar and less intelligent than humans, would quickly exceed the intrinsic novelty of the interaction. The results of this study suggest, however, that this was not the case. In spite of the costs, participants sent a greater number of messages to a chatbot than they did to another human, even though the shorter message lengths and more limited vocabulary in these conversations clearly illustrated the overhead of the adaptive processes present. Alongside these processes, the greater use of profanity in these conversations suggests that participants never lost sight of the fact that they were communicating with a computer.

People obviously have less experience communicating with chatbots than they do with other people. Chatbots are limited in their ability to have an extended goal-directed discussion, and can offer little in the way of common history or shared experience. In spite of these limitations, however, many people are willing to have extensive interactions with chatbots, suggesting that these artificially intelligent systems must be offering enough to capture the attention of millions of users. However, although chatbots seem capable of participating in and holding the attention of human users, natural language interfaces to computer systems cannot yet simulate the full range of intelligent human conversation.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Baron, N. (2007). Text messaging and IM: Linguistic comparison of American college data. *Journal of Language and Social Psychology, 26*(3), 291–298.

Baron, N. (2003). Language of the internet. In A. Farghaly (Ed.), *The Stanford handbook for language engineers* (pp. 59–127). Stanford: CSLI Publications.

Bloom, L., Rocissano, L., & Hood, L. (1976). Adult-child discourse: Developmental interaction between information processing and linguistic knowledge. *Cognitive Psychology, 8*, 521–552.

Carpenter, R. (n.d. a). Jabberwacky. <http://www.jabberwacky.com>.

Carpenter, R. (n.d. b). Cleverbot. <http://www.cleverbot.com>.

Coniam, D. (2008). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL, 20*(1), 98–116.

Craig, D. (2003). Instant messaging: The language of youth literacy. *The Booth Prize Essays*. <https://undergrad.stanford.edu/programs/pwr/publications-prizes-and-awards/boothe-prize-essays>.

Derks, D., Fischer, A. H., & Bos, A. E. R. (2008). The role of emotion in computer-mediated communication: A review. *Computers in Human Behavior, 24*(3), 766–785.

Driscoll, D. (2002). The Ubercool morphology of internet gamers: A linguistic analysis. *Undergraduate Research Journal for the Human Sciences, 1*. <http://www.kon.org/urc/driscoll.html>.

Ferguson, C. A. (1975). Toward a characterization of English foreigner talk. *Anthropological Linguistics, 17*(1), 1–14.

Ferrara, K., Brunner, H., & Whittemore, G. (1991). Interactive written discourse as an emergent register. *Written Communication, 8*, 8–34.

Gianotto, A. (n.d.). BanBuilder banned – Word list generator. <http://www.banbuilder.com>.

Hasler, B. S., Tuchman, P., & Friedman, D. (2013). Virtual research assistants: Replacing human interviewers by automated avatars in virtual worlds. *Computers in Human Behavior, 29*, 1608–1616.

Hausendorf, H. (1992). Patterns of adult-child interaction as a mechanism of discourse acquisition. *Journal of Pragmatics, 17*, 241–259.

Hentschel, E. (1998). Communication on ircLinguistik. *Linguistik Online*. <http://www.linguistik-online.de/irc.htm>.

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

Internet Slang Words. (n.d.). <http://www.internetslang.com>.

Jay, T. (2009). The utility and ubiquity of taboo words. *Perspectives on Psychological Science, 4*(2), 153–161.

Katsuno, H., & Yano, C. R. (2002). Face to face: On-line subjectivity in contemporary Japan. *Asian Studies Review, 26*(2), 205–231.

Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Lortie, C. L., & Guitton, M. J. (2011). Judgment of the humanness of an interlocutor is in the eye of the beholder. *PLoS One, 6*(9), e25085. http://dx.doi.org/10.1371/journal.pone.0025085.

Aron, J. (2011). Software tricks people into thinking it is human. *New Scientist*. <http://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human.html>.

Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). LIWC2007 manual: The development and psychometric properties of LIWC2007. LIWC.net.

Shawar, B. A., & Atwell, E. (2005). Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics, 10*(4), 489–516.

Shawar, B. A., & Atwell, E. (2007). Chatbots: Are they really useful? *LDV-Forum, 22*(1), 29–49.

Tagliamonte, S., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. *American Speech, 83*(1), 3–32.

Varnhagen, C., McFall, P., Pugh, N., Routledge, L., Sumida-MacDonald, H., & Kwong, T. (2009). Lol: New language and spelling in instant messaging. *Reading and Writing, 23*(6), 719–733.

Wallace, R., Tomabechi, H., & Aimless, D. (2003). Chatterbots go native: Considerations for an ecosystem fostering the development of artificial life forms in a human world. <http://www.pandorabots.com/pandora/pics/chatterbotsgonative.doc>.

Weizenbaum, J. (1966). ELIZA: A computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45.

Werry, C. C. (1996). *Linguistic and interactional features of internet relay chat*. Philadelphia, PA: John Benjamins Publishing.

Yates, S. J. (1996). Oral and written linguistic aspects of computer-conferencing. In S. C. Herring (Ed.), *Computer-mediated communication: Linguistic, social and cross-cultural perspectives* (pp. 29–46). Philadelphia, PA: John Benjamins Publishing.

Zabaware. (n.d.). Ultra Hal Assistant. <http://www.zabaware.com>.