12-2014

Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment

Ryan M. Schuetzler
University of Nebraska at Omaha, rschuetzler@unomaha.edu

Mark Grimes
University of Arizona

Justin Scott Giboney
University at Albany, State University of New York

Joesph Buckman
University of Arizona

Facilitating Natural Conversational Agent Interactions: Lessons from a Deception Experiment

Completed Research Paper

Ryan M. Schuetzler
University of Arizona
McClelland Hall, Room 427
1130 E Helen St, PO Box 210108
Tucson, AZ 85721
ryan@schuetzler.net

G. Mark Grimes
University of Arizona
McClelland Hall, Room 427
1130 E Helen St, PO Box 210108
Tucson, AZ 85721
mgrimes@cmi.arizona.edu

Justin Scott Giboney
University at Albany
1400 Washington Ave.
Albany, NY 12222
jgiboney@albany.edu

Joseph Buckman
University of Arizona
1209 E. University Blvd.
Tucson, AZ 85721
jbuckman@email.arizona.edu

Abstract

This study reports the results of a laboratory experiment exploring interactions between humans and a conversational agent. Using the ChatScript language, we created a chat bot that asked participants to describe a series of images. The two objectives of this study were (1) to analyze the impact of dynamic responses on participants' perceptions of the conversational agent, and (2) to explore behavioral changes in interactions with the chat bot (i.e. response latency and pauses) when participants engaged in deception. We discovered that a chat bot that provides adaptive responses based on the participant's input dramatically increases the perceived humanness and engagement of the conversational agent. Deceivers interacting with a dynamic chat bot exhibited consistent response latencies and pause lengths while deceivers with a static chat bot exhibited longer response latencies and pause lengths. These results give new insights on social interactions with computer agents during truthful and deceptive interactions.

Keywords: Deception, chat bot, conversational agent, human-computer interaction

Introduction

Enabling humans to communicate with computers as naturally as they would with another human being has long been a goal of researchers working in the area of human-computer interaction. Facilitating natural communication with computers has a number of benefits, including improved communication and task performance in computer-based tasks (Gong 2008; Qui and Benbasat 2009), increased disclosure of information (Moon 2000) and greater veracity in communication (Lind et al. 2013). To this end, a long line of responsive conversational agents (CA), or chat bots, have captured the imagination of researchers for over five decades.

In the 1960's, chat bots such as ELIZA, the computer psychologist, began to emerge (Weizenbaum 1966). ELIZA and other chat bots of the day had rudimentary abilities to engage in conversations, typically following simple decision trees to determine what to say next. As technology has advanced, so have chat bots, many of which now leverage natural language processing (NLP), semantic networks, and pattern recognition to provide a more natural chat experience (Wallace 2004).

Advances in technology have increasingly enabled developers to give chat bots more features and abilities. Without understanding the positive or negative consequences of introducing new features into chat interactions, however, it is unclear how these features impact the quality of communication between the human and the computer. Thus, there is a significant need to more fully understand how users' perceptions and responses change as chat bot interactions evolve. In this research, we set out to gain a deeper understanding of how the level of interactivity a chat bot exhibits affects the human participant's interaction with the system.

Despite significant research and technological advances, state-of-the-art chat bots still fail to provide convincing human-like interactions. One reason for this shortcoming is that chat bots do not have access to traditional non-verbal cues. In face-to-face communication, over 60% of information in a message is transmitted non-verbally through changes in features such as vocalic patterns, proxemics, and oculometrics (Birdwhistell 1955; Burgoon et al. 2009). Currently chat bots rely primarily on the text that is provided to them, conducting linguistic analysis to infer meaning, devoid of many of the important cues in the message that humans naturally use to help determine meaning. However, chat bots, if so designed, have access to novel data that is not readily available to human observers, such as small changes in response latency (i.e., the time between when the chat bot finishes asking a question and the user responds) and pause time between words and thoughts. By observing changes in these features, we suggest it may be possible for chat bots to infer more information about a human's disposition, thus increasing the quality of the communication.

RQ1: *How does a chat bot that provides dynamic, rather than static, responses influence user perceptions, responses, and behavior in a real-time chat environment?*

While chat bots are currently used for customer service (Gong 2007), entertainment (Isbister 2006), and healthcare (Bickmore and Picard 2005), a novel application of chat bots is conducting credibility assessment interviews. Assessing credibility in computer-interactions has many useful applications. For law enforcement, credibility assessment may be a primary goal. For other applications, such as medical practitioners, credibility assessment may be useful for detecting when patients are withholding information that may impact their diagnosis - for example, underreporting unhealthy behaviors or failing to list all prescription medications could both have substantial effects on medical diagnoses and care. Measuring and improving adherence to treatment instructions is an important topic of research in healthcare (Steiner 2012). Assessing veracity in text-based computer-mediated communication, however, is a difficult and as of yet unresolved challenge (Zhou, Burgoon, Nunamaker, et al. 2004). Due to the importance and difficulty of this issue, a second focus of this research is to explore how chat bot interactions change under conditions of deception. Of the deception cues, in this study we specifically examine response latency and pause length. In accordance with Interpersonal Deception Theory (IDT), deceivers may require additional time to fabricate a message and may exhibit longer pauses compared to non-deceivers (Buller and Burgoon 1996). As such, we seek to answer:

RQ2: *How does a chat bot that provides dynamic, rather than static, responses influence user behavior when engaging in deception?*

In this study, a chat bot CA was built using the ChatScript engine. A two-condition laboratory experiment was conducted in which participants chatted with either a static CA which followed a predetermined script or a dynamic CA which adjusted its responses based on the text provided by the participant. During the course of the conversation, participants viewed images and were asked to describe the images to their chat partner either truthfully or deceptively. In the dynamic chat bot condition, NLP techniques embedded in the ChatScript program were used to formulate a reasonably human-like response, while in the static condition a standard list of questions was asked.

Based on the findings of this research, we discuss implications for practice and propose a way forward for additional research using chat bots for interactive communication, specifically for scenarios in which credibility assessment or elicitation of sensitive information is important.

Literature Review

Significant prior research in CAs, social presence, and deception provides a mature foundation for this research. In the following sections we discuss some of the relevant extant literature that informs this work.

Conversational Agents

The first well-known chat bot, ELIZA, used keyword recognition and substitution when analyzing human input and formulating responses (Gianvecchio et al. 2011; Weizenbaum 1966). It was one of the first computer programs to use rudimentary NLP for responding to users. ELIZA used scripts containing ranked keywords and transformation rules to respond to users based on the text they provided. While the scripts often resulted in rather unengaging conversations, ELIZA was considered very human-like in its time (Weizenbaum 1966). The architecture behind ELIZA proved to be a significant step forward for CAs and paved the way for future chat bot design and creation.

Another significant step forward for chat bots came with the creation of A.L.I.C.E. (Artificial Linguistic Internet Computer Entity). The major contribution of A.L.I.C.E., which is freely available for download at http://www.alicebot.org/downloads, is the introduction of AIML (Artificial Intelligence Markup Language). AIML is a programming language similar to XML that enables the creation of a chat bot using pattern-matching to parse user input. AIML also provides additional features that allow programmers to control the chat bot's behavior, which has led to it becoming a staple programming language for chat bot creation. New variations of AIML that offer their own unique set of features are being developed, For example, the Multimodal Presentation Markup Language (MPML) (Descamps et al. 2001; Mori et al. 2003; Prendinger et al. 2004) provides the ability to express emotional behavior. ChatScript (http://chatscript.sourceforge.net/) is another language, separate from AIML, which provides a similar feature set. These programming languages have enhanced abilities to control the chat bot's interactivity, providing the tools necessary to recognize linguistic features and provide human-like responses.

Due to the prevalence of nonverbal cues in human communication, chat bots sometimes struggle to capture the meaning and context of messages. This poses a significant challenge for a chat bot to communicate effectively with a human. For this reason, a great deal of prior chat bot research focuses on improving the extraction of linguistic features from text (Guo and Zhang 2009; Santangelo et al. 2006; Shawar and Atwell 2004). The process of extracting meaning from text and forming a response is known as natural language processing.

NLP is a theoretical set of computational techniques for analyzing and representing naturally occurring texts at one or more levels of linguistic analysis for the purpose of achieving human-like language processing for a range of tasks or applications (Liddy 2001). Some chat bots use sophisticated linguistic analysis to interpret a human's response. However, lower-level systems do not aim to actually understand the response, but rather mimic understanding through pattern-matching (Schumaker and Chen 2010). NLP gives chat bots the ability to analyze a human's response by searching for keywords. Therefore, part of preparing a simple chat bot is determining what a reasonable set of responses might be in order to eliminate the need for advanced linguistic analysis. A chat bot engages in conversation by recognizing when a keyword appears and using it to formulate the next response or question. If a chat bot does not recognize a keyword in the response it may request clarification or change the subject.

As chat bot techniques evolve to improve chat bot interactivity, applications of chat bots are expanding into new domains such as public health (Crutzen et al. 2011; Johnson and Cooper 2009), tourism (Santangelo et al. 2006; Vlahakis et al. 2001), psychology (Augello et al. 2008; Pilato et al. 2005), education (André 2008; Kerly et al. 2007; Kim et al. 2008; Kowalski et al. 2011), customer support (Brown and Barros 2013; Kuligowska and Lasek 2011), and deception detection (Derrick et al. 2013; Nunamaker Jr et al. 2011). The primary focus of most of these studies has revolved around user acceptance and chat bot functionality. With the exception of Nunamaker et al. (2011), there are few chat bot studies that investigate how the interactivity of a chat bot affects the user's perceptions of the chat bot and their subsequent behavior. Therefore, in this study we investigate how interactive chat bots affect perceptions of social presence and human behavior during deception.

Social Presence

Social presence theory describes the salience of an interpersonal relationship as a function of the salience of the other person in the interaction (Short et al. 1976). Social presence theory is typically applied to interpersonal interactions, though in recent research, and because of our understanding of the paradigm of Computers as Social Actors (CASA) (Nass et al. 1994), the same principles can be applied to interactions with computers (Tourangeau et al. 2003) or with web sites (Gefen and Straub 2004). The CASA paradigm applies the application of interpersonal theories and norms to human-computer interaction.

Social presence theory suggests that the more salient one's partner is in an interaction, the more that person will be focused on managing the relationship with that partner. For example, a video conference makes the other person more salient than does an instant message (IM) conversation. So in an IM conversation, one might pay less attention to managing their relationship with the other person than they would in a video conference or face-to-face. A CA that implements NLP to understand what the user is saying and uses that understanding to make sensible responses to the user will likely increase the feeling of social presence. Compared to a static interview system that asks the same questions regardless of user feedback, we anticipate that a dynamic interview in which the CA asks follow-up questions in response to user messages will increase feelings of social presence. We propose that creating a sense of social presence with a CA will increase the user's perception of the chat bot's humanness and partner engagement. Perceived partner humanness is based on the concepts introduced by Alan Turing (1950) to test whether a computer can convince a person that it is a human and is operationalized here as a rating on a 6-point scale from "definitely computer" to "definitely human" (Ijaz et al. 2011). Perceived partner engagement is how well the computer interacts as a communication partner and is rated on attributes such as skilled, thoughtful, and polite (Holtgraves and Han 2007).

H1: *A CA that adapts responses based on the content of a user's message will result in higher perceived partner humanness*

H2: *A CA that adapts responses based on the content of a user's message will result in higher perceived partner engagement*

Deception

An emerging area of research is investigating changes in human behavior resulting from changes in the CA with which they are engaging (Nunamaker et al. 2011). One area of interest is behavioral changes resulting from the human lying to a CA (Zhou, Burgoon, Nunamaker, et al. 2004; Zhou, Burgoon, Twitchell, et al. 2004; Zhou, Burgoon, Zhang, et al. 2004). This section describes the hypothesized impact of a dynamic chat bot on human behavior compared to a static interviewer by introducing Interpersonal Deception Theory (IDT) (Buller and Burgoon 1996; Burgoon et al. 1999) and explaining how chat bots are predicted to engender strategic behavior described by IDT.

IDT is a communication theory that posits, among other things, that deceivers strategically manage expressed information and behavior more than truth-tellers (Buller and Burgoon 1996). IDT also predicts that as the communication setting becomes more interactive (i.e., more opportunities for each party to react to the other), deceivers will even more greatly manage their expressed information and control their behavior with the objective of appearing more credible. To manage expressed information, deceivers will be more deliberate in their choice of words and actions (Elkins and Stone 2011).

One setting in which deception may occur is in an interview where one communicator asks questions while the other answers the questions. One way of promoting more interactive communication in an interview setting is for the interviewer to ask follow-up questions to the interviewee. Follow-up questions are questions based on information provided by one's communication partner (Levine and McCornack 1996). Although follow-up questions do not necessarily increase accuracy of deception detection (Buller et al. 1989, 1991; Stiff and Miller 1986), they can lead to the deceiver presenting additional cues of deception. Prior research on follow-up questions suggests that during follow-up questions deceivers display more speech errors, talk less, pause more, and wait longer before responding (i.e., increased response latency) (Buller et al. 1989, 1991; Derrick et al. 2013; Levine and McCornack 1996).

Because communication with a CA can have varying degrees of interactivity, IDT would suggest that deceivers communicating with CAs that are more interactive are more likely to strategically manage expressed information and control their behavior. CAs that are more interactive (i.e., ask follow up questions) should increase a deceiver's use of strategic behaviors. Thus, deceivers communicating with a dynamic chat bot will be more deliberate in their communication and will exhibit smaller behavioral changes from their baseline. The behaviors we monitor are response latency and pause length. A truth teller is likely to feel a lesser sense of urgency to respond quickly when a CA is dynamic where a deceiver interacting with a dynamic CA is likely to feel a greater sense of urgency due to their desire to appear truthful. When the CA is not dynamic, a deceiver feels less urgency. Thus, deceivers with a static CA will take a greater amount of time to refine their message and appear truthful. Based on this line of reasoning, we propose the following hypotheses:

H3: A CA that adapts responses to the content of the user's message will lead to deceivers exhibiting response latency closer to their baseline than deceivers in a static interview.

H4: A CA that adapts responses to the content of the user's message will lead to deceivers exhibiting pause lengths closer to their baseline than deceivers in a static interview.

Methodology

We performed a laboratory experiment at a large public university in the southwestern United States with sixty students (44 male) from an upper-level MIS class who participated in exchange for class credit. Technical issues with the chat program invalidated data from three participants. Two participants indicated that they did not follow the instructions (i.e. they did not lie when asked) and were therefore removed from the analysis, leaving 55 participants for the final analysis. The experiment took place in a computer lab equipped with privacy screens to prevent participants from seeing stimuli presented on the screens of other participants. Upon arrival at the lab, participants signed a consent form and completed a computer-based survey measuring demographic information and computer use behavior.

After completing the initial survey, participants were automatically directed to the experiment. The system randomly assigned participants to a dynamic chat bot or a static interview treatment (see Table 1 for the experiment flow). Thus, we designate our first independent variable to be whether the chat bot is dynamic or static. This manipulation was a between-subjects manipulation, as each participant only interacted with a single chat bot. During the experiment, participants were shown a series of twelve images taken from the International Affective Picture System (IAPS) (Lang et al. 2008). IAPS is a collection of over 1,000 images covering a wide range of settings and activities that have been validated across hundreds of studies to elicit specific affective states. The images appeared in the same order to each participant, and alternated between positive and negative valence. While the valence was a within-subjects manipulation, it is not used for analysis in this study. After participants finished describing an image, the next image appeared. For a pre-selected set of three negative images (the same images for all participants), participants were asked to be lie about the content of the image, for example, by respond to questions regarding the image of a sickly dog as if it was a happy dog – a methodology similar to that used by Ekman and Friesen (1974) – making our second independent variable the message's veracity. This was a within-subjects manipulation. Figure 1 is a sample interaction from the deceitful condition.

<table>
<tr><td rowspan="2">Pre-survey</td><td>Random assignment to static interview</td><td>Image treatments (positive and negative)</td><td rowspan="2">Typed responses to questions (truthful and deceptive).
Typing latency and pause time recorded.</td><td rowspan="2">Post-survey</td></tr>
<tr><td>Random assignment to dynamic chat bot</td><td>Image treatments (positive and negative) with dynamic follow-up questions</td></tr>
</table>

Table 1. Experiment design

Before interacting with the chat bot, the participants were given the following instructions:

For this experiment you will interact with either a computer or a human via chat. You will be shown a series of images which your chat partner cannot see. Your chat partner will ask you to describe each image, and may ask for more details on some of the images.

For some images you will be instructed to lie to your chat partner by describing something DIFFERENT than what is in the image.

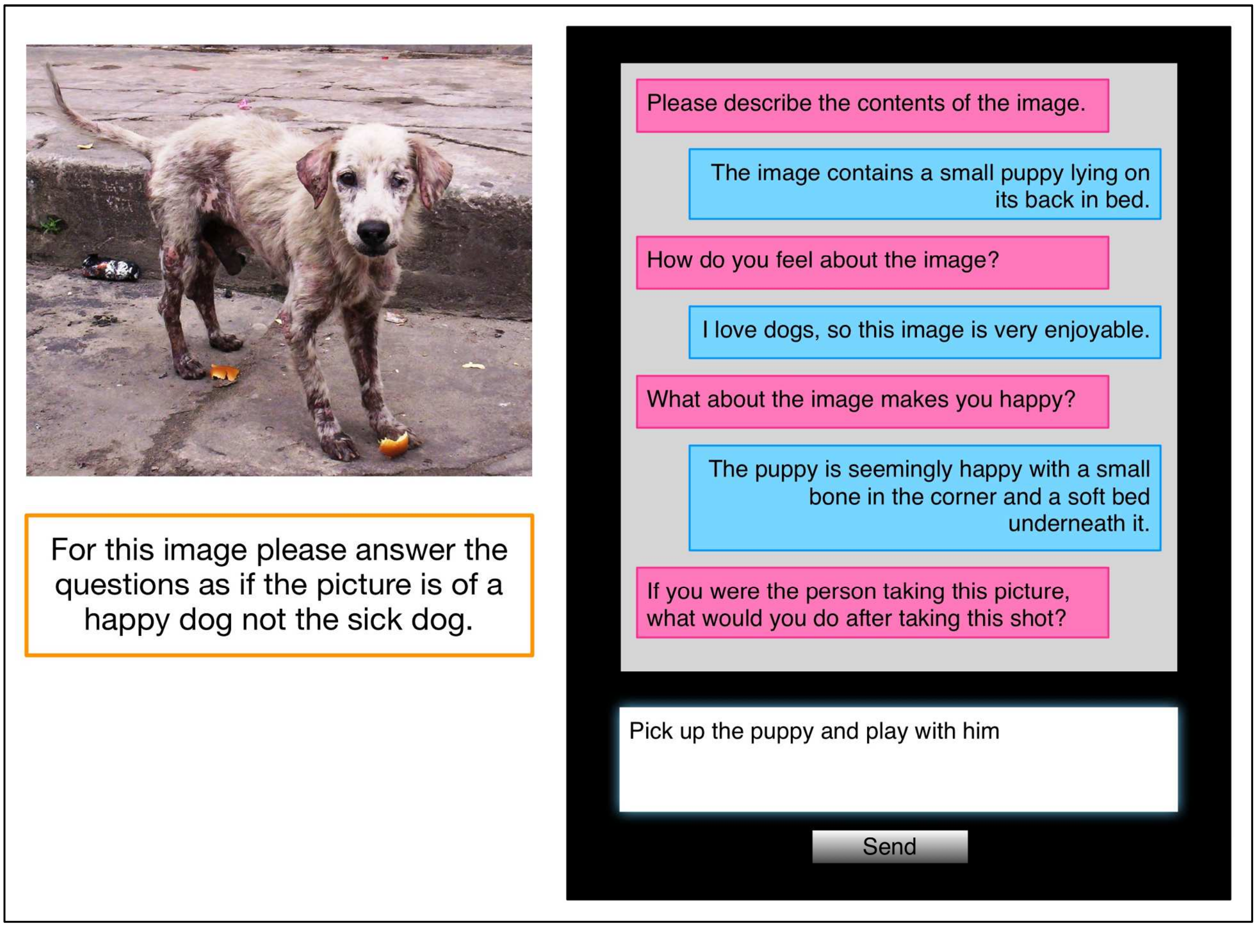

Figure 1. Example chat interface (Pink bubbles are text from the chat bot)[1]

While a human condition is mentioned in the instructions so that participants will be receptive to the idea that their chat partner might be human, no human chat interactions occurred.

Participants were then shown two example screenshots of the interface. In the first screenshot, the participants were shown what the chat interface would look like when they were to respond truthfully, and in the second screenshot they were shown what the interface would look like when they were to behave deceptively - the addition of a bold message stating "For this image, please answer the questions as if the picture is of [something similar but positive] not of [the actual subject of the image]." Neither of the images used in the screenshots were used in the experiment, and both images were of neutral valence (shoes and clothes pins).

After the two screenshots, the user clicked a link that took them to the chat interface (Figure 1). The chat bot followed a conversation stream similar to that presented in Table 2 asking each participant three base questions, and in the dynamic condition, a follow up question to each base question. The follow up

[1] Dog image from https://www.flickr.com/photos/amazoncares/83761580/ under the Creative Commons license https://creativecommons.org/licenses/by/2.0/

questions looked for key words in the responses – for example, as illustrated in Figure 1, since the user said the key word "enjoyable" the chat bot specifically asked "What about the image makes you happy?" See the appendix for examples of additional conversations. As previously described, next to some of the images appeared a prompt instructing participants to lie to their chat partner and describe a different scene. The images where participants were to lie were pre-selected to ensure consistency across participants and conditions. As participants typed descriptions of the images, a small JavaScript application embedded in the web page captured and stored precise measurements of their response latencies and pauses while typing. These measures were used as dependent variables in the testing of H3 and H4.

Static	Dynamic
1. Please describe the contents of the image	1a. Please describe the contents of the image
	1b. [Follow up question based on response]
2. How do you feel about the image?	2a. How do you feel about the image?
	2b. [Follow up question based on response]
3. If you were the person taking this picture, what would you do after taking the shot?	3a. If you were the person taking this picture, what would you do after taking the shot?
	3b. [Follow up question based on response]

Table 2. Static and Dynamic Interview Flow

After answering questions for all twelve images, participants completed a post survey where they answered questions about their partner and the interaction. The post survey measured the humanness and engagement dependent variables. We also checked the participants' compliance with our instructions to lie about some of the images by asking "For some of the images you were asked to respond to the questions as if a different, more pleasant image was on the screen. Did you complete that task as assigned?" We further verified this information by checking chat logs to ensure participants answered questions as directed. We asked about the chat interaction using a Likert-type scale from 1, strongly disagree, to 7, strongly agree on the following dimensions: reasonable, weird, unusual, natural, and comfortable. We also asked about the chat partner in a similar manner on these dimensions: skilled, polite, engaging, responsive, thoughtful, and friendly (Holtgraves and Han 2007).

We also wanted to see if any of the participants thought that the chat bot might be human. To this end, we asked the following: "My chat partner was…" with the following six options: definitely human; probably human; not sure, but guess human; not sure, but guess computer; probably computer; definitely computer (Ijaz et al. 2011).

Analysis and Results

Our research design had one binary, between-subjects condition (dynamic chat bot or static interview) and two binary, within-subjects conditions (pleasant image or negative image; and answer truthfully or answer deceptively), though differences in the valence conditions are not analyzed here.

To test H1, that individuals in the dynamic chat bot condition would rate their partner as more human-like than those in the static interview condition, we conducted a Welch's two-sample *t*-test. A statistically significant difference between conditions was found (t = -3.38, df = 44.63, p = 0.002), as illustrated in Figure 2. Furthermore, participants in the chat bot condition were more likely to express doubt in the nature of their chat partner, with 79.2% of static interview participants saying their chat partner was definitely a computer, while only 41.9% of those in the dynamic chat bot condition saying the same. Finally, four participants (12.9%) in the dynamic chat bot condition indicated they thought it was more likely their chat partner was human than computer, rating the humanness above the midpoint of the 0-5 scale, while no participants in the static condition provided such a rating. As such, H1 is supported.

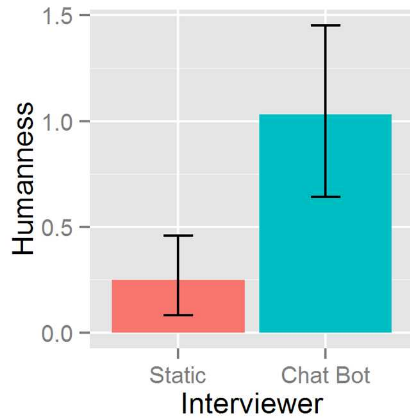

Figure 2. Humanness Rating by Condition (Rated from 0 to 5)

To test H2, that individuals in the dynamic chat bot condition would rate their partner as more engaging than those in the static interview, we conducted another Welch's two-sample *t*-test. The result indicates a statistically significant difference between the two conditions (t = -3.88, df = 40.09, p < 0.001), providing support for H2. Participants in the chat bot condition rated their partner as more engaging on a seven-point scale (M = 5.0, SD = 1.0) than those in the static interview condition (M = 3.7, SD = 1.4). Figure 3 shows the difference between conditions.

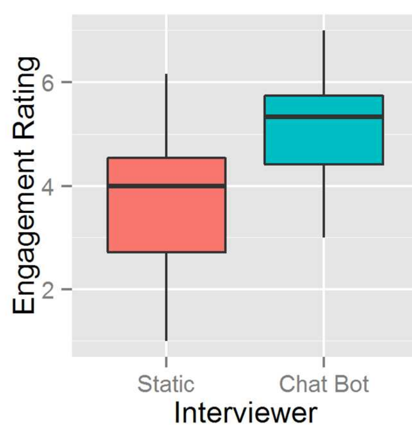

Figure 3. Partner Engagement by Condition (Rated from 1 to 7)

For analysis of H3, we compared response latency between subjects when deceiving and telling the truth. We used a mixed effects model to account for the fact that both response latency and pause time for deceptive and truthful interactions were repeated measures. To compare conditions on response latency, we took each individual's mean response latency for the unpleasant truthful and unpleasant lie images and compared them. Pleasant images were, for these analyses, used as a buffer between unpleasant images. To be consistent across conditions, we used only the response latency for the initial questions, not using follow-up questions for this analysis. In addition, because the distribution of response latency was right-skewed, we used a log transformation, resulting in a more normal distribution.

The result of the mixed model showed a significant interaction between deception and condition (β = -0.22, SE = 0.10, p = 0.03). There was also a statistically significant main effect of deception (β = 0.23, SE = 0.07, p = .002). There was no significant main effect of condition (β = -0.07, SE = 0.10, p = 0.47). This provides support for H3: those in the dynamic chat bot condition showed greater consistency when lying than did those in the static interview condition. This result is illustrated in Figure 4, with details of the model comparison in Table 3. During truthful questions, response latency was approximately equal between conditions, but when lying those in the static interview condition showed increases in latency while the dynamic chat bot condition did not.

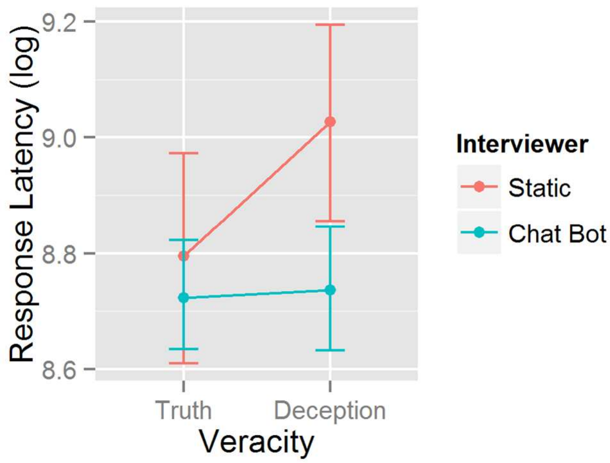

Figure 4. Means Plot of Response Latency

Variable	Model 1 B	Model 1 SE	Model 2 B	Model 2 SE	Model 3 B	Model 3 SE	Model 4 B	Model 4 SE
(Intercept)	8.81***	0.04	8.75***	0.05	8.86***	0.07	8.80***	0.07
D			0.11*	0.05	0.11*	0.05	0.23**	0.07
R					-0.18*	0.09	-0.07	0.10
D x R							-0.22*	0.10
-2LL	81.3		76.6		72.2		67.1	
L Ratio	—		4.72*		4.39*		5.12*	

Notes: *N*=55. * = *p* < .05, ** = *p* < .01, *** = *p* < 0.001. D = deception (within subjects); R = responsiveness (between subjects). -*2LL* is a measure of model fit that, when compared using the likelihood ratio test (L Ratio), approximates a chi-square distribution and is therefore useful in comparing nested models.

Table 3. Summary of Linear Mixed-Effects Models for Response Latency

To test H4, we used the mean of pauses during the response. We began by removing all pauses shorter than 500 ms, as those were deemed to be more indicative of normal latencies in typing rather than of pauses for thinking. We then took the mean of those pause times for both unpleasant truth and unpleasant deception questions. As this distribution was also right-skewed, we used a log transformation to produce a more normal distribution.

Results of the mixed model looked very similar to the response latency analysis for H3. Once again, we find a significant main effect for deception ($\beta = 0.18$, $SE = 0.41$, $p < 0.001$), but none for condition ($\beta = -0.09$, $SE = 0.08$, $p = 0.26$). Most importantly, the interaction effect of deception and condition was statistically significant ($\beta = -0.17$, $SE = 0.06$, $p = 0.003$). As in the response latency analysis, the interaction effect is approximately equal and opposite the deception main effect, meaning that deception increased mean pause length, but only for those in the static interview condition (Figure 5). Details of the model comparison are shown in Table 4.

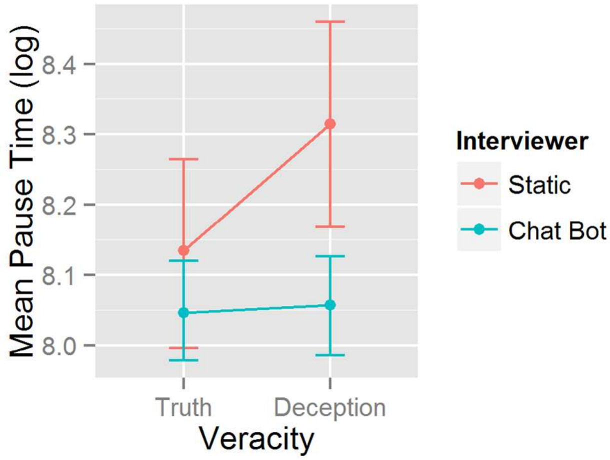

Figure 5. Means Plot of Mean Pause Length

	Model 1		Model 2		Model 3		Model 4	
Variable	***B***	***SE***	***B***	***SE***	***B***	***SE***	***B***	***SE***
(Intercept)	8.13***	0.04	8.08***	0.04	8.18***	0.06	8.13***	0.06
D			0.08**	0.03	0.08*	0.03	0.18***	0.04
R					-0.17*	0.07	-0.09	0.08
D x R							-0.17**	0.06
-2LL	9.9		2.1		-3.5		-12.5	
L Ratio	—		7.81**		5.60*		9.04**	

Notes: *N*=55. * = *p* < .05, ** = *p* < .01, *** = *p* < 0.001. D = deception (within subjects); R = responsiveness (between subjects). -*2LL* is a measure of model fit that, when compared using the likelihood ratio test (L Ratio), approximates a chi-square distribution and is therefore useful in comparing nested models.

Table 4. Summary of Linear Mixed-Effects Models for Mean Pause Length

Discussion

The results of H1 highlight how trivial it is to increase the perception of humanness of a chat bot. While the current chat bot did leverage NLP techniques to craft answers that were relevant to the conversation at hand, the responses were still very basic in nature. The appendix shows a few examples of the simplicity of our chat bot's responses. Regardless, our rudimentary chat bot was able to fool four users and at least introduce doubt into many others. Given the limited nature of the chat interaction, it is intriguing that we were able to introduce humanness to our computer system. We suspect this result is likely similar to what might happen if a static interview was compared to an interview with human agent in this text-based scenario – an area that will be explored in more depth in future research.

The results of H1 and H2 provide evidence of the social presence effect of adding interactive and responsive communication to a CA. Large changes in user perceptions of the system came from small cues indicating the chat bot understood users. Further refinement of the current chat agent could be done using the chat logs to understand areas where our chat agent failed to match user input. Iterative development such as this would likely enable even greater gains in ratings of humanness and engagement.

The support for hypotheses 1 and 2 is interesting in its own right, but even more interesting as it relates to H3 and H4. We found that when people interact with a dynamic chat bot as opposed to a static interviewer, they perceived their chat partner as both more engaging and more human-like. How people interact with a CA is likely to be affected by that change. The CASA paradigm predicts that people will view computers as social agents and treat them as such, applying social norms and expectations, even when they do not think of the system as explicitly social (Nass et al. 1994).

We suggest that the use of these social norms drove the support of H3 and H4. In the static interview condition, participants felt they could take all the time they needed to concoct a deceptive response when asked. Contrast that with the participants who were assigned a dynamic chat agent. Presumably because they felt their partner was more engaging and more human-like, they reciprocated by behaving in a more engaging way. Part of behaving in an engaging way included maintaining their normal response latency and pause times. These results make sense if we imagine them in the context of interpersonal communication. For example, one has a normal pattern of communication when interacting with a friend, which consists of specific response latencies, pausing to think, etc. If we were lying during this communication, interpersonal deception theory posits that we will try to maintain our normal pattern of behavior, including not taking extra time to create our response (Buller and Burgoon 1996). Contrast that with lying to a system that is unresponsive to our behavior, as in the static interview condition of this study, if the system gives no indication of interpreting or understanding our response, individuals may be more likely to take as much time as they feel is necessary to craft a response.

These findings have interesting implications for practitioners. First, we see that making modest efforts toward making a system more human-like can reap large rewards with regard to user perceptions of the system. Second, by monitoring features such as response latency and pause times, we may be able to identify when users are behaving deceptively. Further, we found that this ability may be enhanced by using CAs that are less engaging – a finding that is somewhat counterintuitive and deserves further investigation. Finally, this research suggests that for eliciting information, creating a more human-like ECA may be useful so that interviewees, attempting to maintain reciprocity in the communication by responding expeditiously, may have less time to carefully construct a response.

Limitations and Future Research

As with all research, this study has limitations, and many of those limitations provide opportunities for future research. First, response latency and pauses are not the only measures of deception that could possibly be affected by the introduction of a dynamic chat bot. Linguistic features and keyboard behavior are just a few of the features that could potentially be examined in future research.

As a first look at the influence of chat bots on deceptive behavior, this study was further limited by the lack of motivation for the deceivers. They were simply told to lie, with no particular incentive other than the instruction given to them. Future research should examine the impact of motivation, or different types of deception. If the chat bot is conducting an interrogation rather than a simple Q&A, there is the potential that the humanness factor becomes an even greater driver.

This study also provides an opportunity for future research in voice-based communication. The current research was text-only, which provides limited cues compared to the voice. This limitation was partially imposed by technology. Because text-to-speech and speech-to-text technology are currently highly active areas of research, the capabilities of these systems will improve rapidly over the coming years. Further, the addition of a face and voice to the intelligent agent could provide an even greater sense of social interaction, causing even greater social responses.

There were only 55 participants in this study in two between-subject conditions. This is a low number of participants for a typical IS study leading to possible type 2 errors. Our student based subject pool also limits the generalizability of our results, however, as an exploration of potential effects of CAs, we believe this still provides a good starting point for future research.

Finally, because this research is designed to compare social responses, future research might compare the results of a chat bot with results of a real human in a chat interaction. Providing a human interaction partner and comparing results to a dynamic chat bot could test the idea that the effects are indeed social in nature.

Conclusion

Our research found that people perceive a dynamic chat bot to be both more engaging and more human-like than a static interview. While this is an intuitive finding, this study helps to quantify the impact of a modest improvement in chat bot interactivity and paves the way for future research to compare how much more human-like and engaging chat bots are as incremental features are added. This research also reveals a less intuitive finding that is useful for practitioners looking to elicit information from subjects. While many theories of deception suggest that deceivers should exhibit additional response latency as they work to craft a response, we found that, in line with IDT, deceivers engaged in more strategic behaviors by responding with lower latency in order to appear engaged to their chat partner. By having more human-like interview agents, deceivers may feel they have less time to formulate a deceptive response, and therefore may be more likely to provide cues to their deception. This shows both the potential benefits of interactive chat agents and the possible downsides. Deception is a complicated behavior, and difficult to detect in text-based communication. The introduction of social agents such as chat bots, if not properly implemented, may make that process even more difficult. Finally, this study provides support for the efficacy of chat bots for new interviewing applications such as medical and law enforcement and provides a starting point for future research in the area.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 1068026. We also gratefully acknowledge the support and guidance of Drs. Judee Burgoon and Jay Nunamaker.

References

André, E. 2008. "Design and Evaluation of Embodied Conversational Agents for Educational and Advisory Software," in *Handbook of Conversation Design for Instructional Applications*, Hershey, PA: Information Science Reference, pp. 668–670.

Augello, A., Saccone, G., Gaglio, S., and Pilato, G. 2008. "Humorist Bot: Bringing Computational Humour in a Chat-Bot System," in *International Conference on Complex, Intelligent, and Software Intensive Systems*, F. Xhafa and L. Barolli (eds.), Piscataway, NJ: IEEE Computer Society, March, pp. 703–708.

Bickmore, T., and Picard, R. W. 2005. "Establishing and Maintaining Long-Term Human- Computer Relationships," *ACM Transactions on Computer-Human Interaction* (12:2), pp. 293–327.

Birdwhistell, R. L. 1955. "Background to Kinesics, Etc.," *A Review of General Semantics* (13:1), pp. 10–28.

Brown, R., and Barros, A. 2013. "Towards a Service Framework for Remote Sales Support via Augmented Reality," in *Workshop on Information Systems and Economics*, pp. 335–347.

Buller, D. B., and Burgoon, J. K. 1996. "Interpersonal Deception Theory," *Communication Theory* (6:3), pp. 203–242.

Buller, D. B., Comstock, J., Aune, R. K., and Strzyzewski, K. D. 1989. "The Effect of Probing on Deceivers and Truthtellers," *Journal of Nonverbal Behavior* (13:3), pp. 155–170.

Buller, D. B., Strzyzewski, K. D., and Comstock, J. 1991. "Interpersonal Deception: I. Deceivers' Reactions to Receivers' Suspicions and Probing," *Communication Monographs* (58), pp. 2–24.

Burgoon, J. K., Buller, D. B., White, C. H., Afifi, W., and Buslig, A. L. S. 1999. "The Role of Conversational Involvement in Deceptive Interpersonal Interactions," *Personality and Social Psychology Bulletin* (25:6), pp. 669–686.

Burgoon, J. K., Guerrero, L., and Floyd, K. 2009. *Nonverbal Communication*, Pearson.

Crutzen, R., Peters, G.-J. Y., Portugal, S. D., Fisser, E. M., and Grolleman, J. J. 2011. "An Artificially Intelligent Chat Agent That Answers Adolescents' Questions Related to Sex, Drugs, and Alcohol: An Exploratory Study," *The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine* (48:5), pp. 514–519.

Derrick, D. C., Meservy, T. O., Jenkins, J. L., Burgoon, J. K., and Nunamaker Jr, J. F. 2013. "Detecting Deceptive Chat-Based Communication Using Typing Behavior and Message Cues," *ACM Transactions on Management Information Systems* (4:2), pp. 1–21.

Descamps, S., Prendinger, H., and Ishizuka, M. 2001. "A Multimodal Presentation Mark-up Language for Enhanced Affective Presentation," in *Proceedings of the International Conference on Intelligent Multimedia and Distant Education*, pp. 9–16.

Ekman, P., and Friesen, W. V. 1974. "Detecting deception from the body or face," *Journal of Personality and Social Psychology* (29:3), pp. 288–298.

Elkins, A. C., and Stone, J. 2011. "The Effect of Cognitive Dissonance on Argument Language and Vocalics," in *Report of the HICSS-44 Credibility Assessment and Information Quality in Government and Business*, M. Jensen, T. O. Meservy, J. K. Burgoon, and J. F. Nunamaker (eds.), pp. 1–8.

Gefen, D., and Straub, D. W. 2004. "Consumer Trust in B2C E-Commerce and the Importance of Social Presence," *Omega* (32:6), pp. 407–424.

Gianvecchio, S., Xie, M., Wu, Z., and Wang, H. 2011. "Humans and Bots in Internet Chat: Measurement, Analysis, and Automated Classification," *IEEE/ACM Transactions on Networking* (19:5), pp. 1557–1571.

Gong, L. 2007. "Is Happy Better Than Sad Even If They Are Both Non-Adaptive? Effects of Emotional Expressions of Talking-Head Interface Agents," *International Journal of Human-Computer Studies* (65:3), pp. 183–191.

Gong, L. 2008. "How Social Is Social Responses to Computers? The Function of the Degree of Anthropomorphism in Computer Representations," *Computers in Human Behavior* (24:4), pp. 1494–1509.

Guo, Q., and Zhang, M. 2009. "Semantic Information Integration and Question Answering Based on Pervasive Agent Ontology," *Expert Systems with Applications* (36:6), pp. 10068–10077.

Holtgraves, T., and Han, T.-L. 2007. "A Procedure For Studying Online Conversational Processing Using a Chat Bot," *Behavior Research Methods* (39:1), pp. 156–163.

Ijaz, K., Bogdanovych, A., and Simoff, S. 2011. "Enhancing The Believability of Embodied Conversational Agents Through Environment-, Self- and Interaction-Awareness," in *Proceedings of the Thirty-Fourth Australasian Computer Science Conference*, Perth, Australia, pp. 107–116.

Isbister, K. 2006. *Better Game Characters by Design: A Psychological Approach*, San Francisco, CA, USA: Kaufmann Publishers.

Johnson, N. A., and Cooper, R. B. 2009. "Media, Affect, Concession, and Agreement in Negotiation: IM Versus Telephone," *Decision Support Systems* (46:3), pp. 673–684.

Kerly, A., Hall, P., and Bull, S. 2007. "Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models," *Knowledge-Based Systems* (20:2), pp. 177–185.

Kim, H., Ruiz, M. E., and Peterson, L. 2008. "Usability and Effectiveness Evaluation of a Course-Advising Chat Bot," *Proceedings of the American Society for Information Science and Technology* (44:1), pp. 1–5.

Kowalski, S., Hoffmann, R., Jain, R., and Mumtaz, M. 2011. "Universities Services in the New Social Eco-systems: Using Conversational Agents to Help Teach Information Security Risk Analysis," in *SOTICS 2011, The First International Conference on Social Eco-Informatics*, , pp. 91–94.

Kuligowska, K., and Lasek, M. 2011. "Virtual Assistants Support Customer Relations and Business Processes," in *Proceedings of the 10th International Conference on Information Management*, Gdańsk.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. 2008. "International Affective Picture System (IAPS): Affective Ratings of Pictures and Instruction Manual," Technical Report No. A-8, Gainesville, FL.

Levine, T. R., and McCornack, S. A. 1996. "A Critical Analysis of the Behavioral Adaptation Explanation of the Probing Effect," *Human Communication Research* (22:4), pp. 575–588.

Liddy, E. D. 2001. "Natural Language Processing," in *Encyclopedia of Library and Information Science*, (2nd ed.) New York: Marcel Decker.

Lind, L. H., Schober, M. F., Conrad, F. G., and Reichert, H. 2013. "Why Do Survey Respondents Disclose More When Computers Ask the Questions?" *Public Opinion Quarterly* (77:4), pp. 888–935.

Moon, Y. 2000. "Intimate Exchanges: Using Computers to Elicit Self-Disclosure from Consumers," *Journal of Consumer Research* (26:4), pp. 323–339.

Mori, K., Jatowt, A., and Ishizuka, M. 2003. "Enhancing Conversational Flexibility in Multimodal Interactions with Embodied Lifelike Agent," in *Proceedings of the International Conference on Intelligent Multimedia and Distant Education*, IUI '03, New York, NY, USA: ACM, pp. 270–272.

Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers are Social Actors," in *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, New York, New York, USA: ACM Press, pp. 72–78.

Nunamaker Jr, J. F., Derrick, D. C., Elkins, A. C., Burgoon, J. K., and Patton, M. W. 2011. "Embodied Conversational Agent-Based Kiosk for Automated Interviewing," *Journal of Management Information Systems* (28:1), pp. 17–48.

Pilato, G., Vassallo, G., Augello, A., Vasile, M., and Gaglio, S. 2005. "Expert Chat-Bots for Cultural Heritage," *Intelligenza Artificiale* (2:2), pp. 25–31.

Prendinger, H., Descamps, S., and Ishizuka, M. 2004. "MPML: A Markup Language for Controlling the Behavior of Life-Like Characters," *Journal of Visual Languages & Computing* (15:2), pp. 183–203.

Qui, L., and Benbasat, I. 2009. "Evaluating Anthropomorphic Product Recommendation Agents: A Social Relationship Perspective to Designing Information Systems," *Journal of Management Information Systems* (25:4), pp. 145–182.

Santangelo, A., Augello, A., and Gentile, A. 2006. "A Chat-Bot based Multimodal Virtual Guide for Cultural Heritage Tours," in *Proceedings of the International Conference on Pervasive Systems & Computing*, Pisa, Italy, pp. 114–120.

Schumaker, R. P., and Chen, H. 2010. "Interaction Analysis of the ALICE Chatterbot: A Two-Study Investigation of Dialog and Domain Questioning," *IEEE Transactions on Systems, Man, and Cybernetics–Part A: Systems and Humans* (40:1), pp. 40–51.

Shawar, B. A., and Atwell, E. 2004. "Accessing an Information System by Chatting," in *Natural Language Processing and Information Systems*, Lecture Notes in Computer Science, Berlin: Springer, pp. 407–412.

Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*, London: Wiley.

Steiner, J. F. 2012. "Rethinking Adherence," *Annals of Internal Medicine* (157:8), pp. 580–585.

Stiff, J. B., and Miller, G. R. 1986. "Come to Think of It...," *Human Communication Research* (12:3), pp. 339–357.

Tourangeau, R., Couper, M. P., and Steiger, D. M. 2003. "Humanizing Self-Administered Surveys: Experiments on Social Presence in Web and IVR surveys," *Computers in Human Behavior* (19:1), pp. 1–24.

Turing, A. M. 1950. "Computing machinery and intelligence," *Mind* (59:236), pp. 433–460.

Vlahakis, V., Karigiannis, J., Tsotros, M., Gounaris, M., and Almeida, L. 2001. "ARCHEOGUIDE: First Results of an Augmented Reality, Mobile Computing System in Cultural Heritage Sites," in *International Symposium on Virtual Reality, Archaeology, and Cultural Heritage*, D. B. Arnold, A. Chalmers, and D. W. Fellner (eds.), Glyfada, Greece, pp. 131–140.

Wallace, R. 2004. *The Anatomy of A.L.I.C.E.*, A.L.I.C.E. Artificial Intelligence Foundation.

Weizenbaum, J. 1966. "Man & Machine," *Computational Linguistics* (9:1), pp. 36–45.

Zhou, L., Burgoon, J. K., Nunamaker, J. F., and Twitchell, D. 2004. "Automating Linguistics-Based Cues for Detecting Deception in Text-Based Asynchronous Computer-Mediated Communications," *Group Decision and Negotiation* (13:1), pp. 81–106.

Zhou, L., Burgoon, J. K., Twitchell, D., Qin, T., and Nunamaker Jr, J. F. 2004. "A Comparison of Classification Methods for Predicting Deception in Computer-Mediated Communication," *Journal of Management Information Systems* (20:4), pp. 139–165.

Zhou, L., Burgoon, J. K., Zhang, D., and Nunamaker, J. F. 2004. "Language Dominance in Interpersonal Deception in Computer-Mediated Communication," *Computers in Human Behavior* (20:3), pp. 381–402.

Appendix – Sample Chat Logs

These are unmodified chat logs taken from participants in the dynamic chat bot condition. These are reproduced verbatim, and any typos were present in the original logs.

User Describing an Image of a Snake

Chatbot: Please describe the contents of the image.
User: a snake about to eat
Chatbot: Tell me more about the snake.
User: it has its mouth open
Chatbot: How do you feel about the image?
User: it is scary
Chatbot: What is scary about it?
User: the snake is attacking something
Chatbot: If you were the person taking this picture, what would you do after taking the shot?
User: run
Chatbot: It does look pretty scary. How far would you have to get away to feel safe?
User: at least 20 feet from it

User Describing an Image of Puppies

Chatbot: Please describe the contents of the image.
User: Three of the cutest black and white puppies ever
Chatbot: Describe the puppies in more detail. What are they doing? What do they look like?
User: They are sitting very well behaved next to each other. The middle one seems to be more excited about the picture but none the less the others are good sports
Chatbot: How do you feel about the image?
User: Happy happy happy
Chatbot: I'm glad this is a pleasant image. What about it makes you happy?
User: the puppies
Chatbot: If you were the person taking this picture, what would you do after taking the shot?
User: grab the puppies and hold them
Chatbot: What else?
User: squeeze them

User Describing an Image of a Smiling Baby

Chatbot: Please describe the contents of the image.
User: This is a picture of a baby in a car seat with his mouth open, sort of smiling.
Chatbot: Why do you think the baby is smiling so much?
User: He isn't smiling that much. but I would assume there is someone behind the camera making him smile like that.
Chatbot: How do you feel about the image?
User: I like this image. Children, babies in particular, are fun to play with and have contagious smiles.
Chatbot: What about the image makes you feel that way?
User: The fact that it is a smiling baby.
Chatbot: If you were the person taking this picture, what would you do after taking the shot?
User: I would probably put the camera don and continue playing with the baby.
Chatbot: What kinds of games do you like to play with babies?
User: Peek-a-boo is a favorite. Also, babies love to be tickled.