

## An Empirical Study Exploring Difference in Trust of Perceived Human and Intelligent System Partners

Joel S. Elson  
University of Nebraska at Omaha  
[jselson@unomaha.edu](mailto:jselson@unomaha.edu)

Douglas C. Derrick  
University of Nebraska at Omaha  
[dcderrick@unomaha.edu](mailto:dcderrick@unomaha.edu)

Luis A. Merino  
University of Nebraska at Omaha  
[lmerino@unomaha.edu](mailto:lmerino@unomaha.edu)

### Abstract

*Intelligent systems are increasingly relied on as partners used to make decisions in business contexts. With advances in artificial intelligence technology and system interfaces, it is increasingly difficult to distinguish these system partners from their human counterparts. Understanding the role of perceived humanness and its impact on trust in these situations is important as trust is widely recognized as critical to system adoption and effective collaboration. We conducted an exploratory study involving individuals collaborating with an intelligent system partner to make several critical decisions. Measured trust levels and survey responses were analyzed. Results suggest that greater trust is experienced when the partner is perceived to be human. Additionally, the attribution of partners possessing expert knowledge drove perceptions of humanness. Partners viewed to adhere to strict syntactical requirements, displaying quick response times, having unnatural conversational tone, and unrealistic availability contributed to perceptions of partners being machine-like.*

### 1. Introduction

Advances in artificial intelligence (AI) and the integration of intelligent systems into the workplace promise to fundamentally change the nature of collaborative work, transforming the way we perceive and relate to computers. Increasingly, individuals are partnering with intelligent systems to enhance productivity and make decisions [5]. These systems may utilize existing interfaces or interact with human users through technology historically reserved for human-to-human communication such as chat and email. The combination of intelligence, capability, natural language, and human-like interaction modalities blurs our ability to discern the identity (human or machine) of such collaborative partners. Understanding the nature of trust in these situations is

essential and pertinent to both system designers and managers responsible for training employees to be effective when utilizing these systems.

Trust can be defined as “the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party” [17]. Trust is widely recognized as being critical in both human-to-human and human-to-machine partnerships. Additionally, trust is an important factor driving the adoption and use of information systems [18] and pivotal for effective and healthy interpersonal relationships between people in the workplace [14, 22]. Trust research has looked extensively into both how individuals trust other humans and how individuals trust technology, yet a comprehensive examination of trust in situations where the identity of the partner is unclear has yet to be conducted.

Contemplating trust in an intelligent system partner perceived as either human or machine-like begs several interesting questions relating to how these types of systems should be integrated into the workplace. For example, should systems be designed to more effectively mask their non-human identities? In what cases should systems be transparent about their intelligent system nature? What cues influence and drive perceptions of a system as being human or machine-like in the first place? And what, if any, effect do these perceptions have on trust? In this work we set out to begin answering these questions. Accordingly, this study has the following research objective:

*To explore the role of perceived humanness of an intelligent system partner and its impact on trust.*

To do this, we conducted a single condition study involving individuals collaborating with an intelligent system partner with an unspecified human-machine identity, and tasked with completing a series of critical decision-making simulations that involved the

perception of real personal risk. At the conclusion of the interaction, individuals self-selected into one of two groups depending on their perception of the partner (partner perceived as human and a partner perceived as chatbot). Trust was measured using a verified instrument and participants were asked to provide open ended responses to questions that intended to capture insight behind the perceptions of their partner. Results suggest that greater trust is experienced when the partner is perceived to be human. Additionally, the attribution of the partner possessing expert knowledge drove perceptions of humanness. A partner viewed to adhere to strict syntactical requirements, displaying quick response times, having unnatural conversational tone, and unrealistic availability contributed to its perceptions of being machine-like.

## 2. Background

In this section we provide a brief overview of trust focusing on interpersonal and swift trust conceptualizations. We highlight important concepts relating to trust in virtual work environments. Finally, we discuss the application of trust to machines and other technology, emphasizing areas of similarity and difference.

### 2.1 A brief overview of trust

Trust is a multidimensional construct with numerous definitions and conceptualizations that vary across domains. Trust in the context of collaborative work (i.e. teams and partnerships) is generally recognized to be a psychological state that is a function of a trustor's individual characteristics (e.g. propensity to trust) along with assessments of a trustee's characteristics (e.g. ability, benevolence, and integrity) and situational factors (e.g. risk, uncertainty). When the level of analysis focuses on trust between individuals, as opposed to trust residing at the team or organizational level, it is referred to as interpersonal trust. In this conceptualization of trust, the individual being trusted (the object of trust) is referred to as the "trustee," while the individual trusting is referred to as the "trustor." Interpersonal trust is important in the development and maintenance of relationships and in the context of work, it has been observed to predict relationship satisfaction as well as overall performance [3].

Another type of trust is swift trust. Swift trust is the initial trust in a teammate or individual trustee in situations where a trustor is required to begin working with another immediately and without ample time to develop trust more gradually [5]. In swift trust,

expectations are applied from more familiar situations along with role expectations when traditional sources of trust are absent.

A triadic framework of trust determinants has been helpful in organizing these sources influencing trust. This framework can be traced back to Kim Giffin, a University of Kansas researcher who published a series of influential papers on trust in the 1960's. The framework consists of three classes of trust determinants: those related to the trustee, those related to a trustor, and ones related to the situation or context where trust is being applied.

Trust determinants related to the trustee may include individual differences and dispositions. Personality traits for example are an important influence on trust, especially early on in trusting relationships [6]. Similarly, one's propensity to trust is an important factor widely recognized to influence initial trust. McKnight, Cummings and Chervany [19] highlighted the importance of an individual's disposition to trust, an individual difference that causes people to trust more or less across situations and trustees.

Important trust determinants related to the trustee include various perceptions of trustworthiness. Foundational work conducted by Mayer, Davis, & Schoorman [17] identified three dimensions at the heart of interpersonal trust including: 1) ability, 2) benevolence, and 3) integrity. Ability describes how capable or skilled a trustee is in carrying out a task in a domain specified by a trustor. Benevolence relates to a trustee having goals or intentions that benefit or align with a trustor. Finally, integrity relates to a trustor and trustee sharing a similar set of values and can be counted on to act in accordance with these shared beliefs.

Determinants related to the situation where trust is being applied include things like interdependence and risk. Interdependence can be thought of as a relationship "where the interests of one party cannot be achieved without reliance upon another" [24]. When collaborating in situations where individuals possess unique skills or abilities necessary for task completion, levels of interdependence may be higher than in situations where individuals are more well versed. Closely related to interdependence is risk, which Deutsch [4] stated was a requisite to trust. Risk has been argued to be a necessary situational factor for trust, with a trustor needing to perceive at least a minimal amount of risk.

In this brief overview of trust, one final aspect needing to be mentioned is the recognition that trust is not static and is influenced by these trust determinants over time. Rousseau et al. [24] emphasize the dynamic nature of trust and outline three phases generally

associated with trust: building, stability, and dissolution. While trust can be said to develop over time, it can also degrade with time. The spiral nature of trust was first described by Golembiewski and McConkie [9] who outlined the self-heightening spirals that exemplify the relationship between trust and risk-taking behavior. Individuals who trust initially will be able to engage in risk-taking behavior which can be rewarded and build trust, which if successfully repeated may continue to increase trust. In the opposite fashion, individuals who distrust initially may not engage in risk-seeking behaviors and reduce or eliminate opportunities to experience positive feedback and thus continue in a spiral of distrust. Initial interactions and experiences are therefore critical to trust development, as early trusting beliefs drive behavior that can build or detract from trust formation.

## 2.2 Trust in people doing virtual work

In studies looking at trust in both in-person and virtual teams, it has been shown that trust starts lower and takes longer to develop in virtual teams than it does in face-to-face teams [13]. The development of trust in virtual work environments may limit the ability of a trustor to gather information related to the trustee and change dynamics of collaboration due to constraints of various technologies.

Early on in an interaction, and with little information available to inform trusting perceptions, disposition to trust is important especially in virtual work contexts [11]. The importance of trusting disposition however may decrease over time as the interaction affords individuals an opportunity to observe one another's ability, integrity, and benevolence. In addition, to trusting disposition, social categorization exerts influence on initial trust in virtual teams. Social categorization is the assignment of partners and teammates into social categories such as gender [27]. Similarity-attraction theory posits that individuals are more likely to be attracted to and develop trust more quickly with individuals who are perceived to be similar to themselves [2]. It follows therefore that the assignment of a virtual partner to a group in which a trustor identifies with themselves would lead to greater initial trust perceptions.

Trust determinants (ability, benevolence, and integrity) related to perceptions of a trustee have also been shown to play a role in trust in virtual collaborative work environments. While initially trust in partners performing computer-mediated work may be lower than in face-to-face interactions, overtime trust levels can reach comparable levels [3]. Key determinants of trust in virtual teams include

perceptions of partner integrity and ability, especially early on in a trusting relationship [11].

## 2.3 Trust in technology

While trust has traditionally been viewed as existing between people, the rise of various automated systems motivated the application of trust to relationships between human and machine. Trust applied to this context is largely based on the view that computers are social actors. Computers as Social Actors Theory is based on empirical evidence suggesting that humans treat computers in many of the same ways as they do other people [21]. Seminal research performed by Clifford Nass, Jonathan Steuer, and Ellen Tauber [21] investigated this phenomenon with a series of experiments and found strong support for computers as social actors. With people viewing computers as social agents, many of the same phenomena found in human interpersonal trust apply to computers and automation. For example, there may be differences of trust between computers perceived to be male or female in ways that are similar to differences in trust between male and female humans [12, 2].

The elements of human-to-human trust are similar in many respects to those that serve as the basis for human-machine trust. Lee and See [13] link three of the dimensions of trust identified by Lee and Moray [12] (performance, process, and purpose) with elements of trust identified in a number of prior human-to-human trust studies including Mayer, Davis & Shoorman [17]. In doing so they highlight that performance relates to ability, process relates to dependability or integrity, and purpose relates to benevolence.

Although there are many similarities between human-to-human and human-machine trust, there do exist a number of differences. There are some researchers who doubt whether or not humans can trust anything but people, let alone computers or objects as these non-living things do not possess will or volition [7]. They define trust as "accepted vulnerability to another" who as the trustee, is assumed to have a lack of good will toward the trustor. With advances in artificial intelligence, the argument for whether machines can possess volition is largely philosophical relating to the nature of intelligences and whether synthetic intelligence can parallel natural intelligence in areas like free will.

More concrete differences between trust in humans and machines have been reported. Pioneering research work in automation found that humans tend to begin the interaction with these systems starting with an initial level of distrust in the system [26].

Foundational work in human-to-human trust found the opposite phenomenon to be true. In human-to-human trust, when an individual first encounters another individual they are initially trusting of that person [15]. A possible reason for this initial propensity to trust in human-to-human relationships was that it reduced the complexity of a new relationship allowing individuals to focus their attention on more important cognitive processes, however this explanation was not able to explain the differences in attitudes of trust toward all computer systems. Additional differences between human-to-human trust and human-machine trust were described in a review of literature focusing on these two types of trust conducted by Madhavan and Wiegmann [16]. Their proposed framework comparing trust development in these two contexts shows a number of areas where human-to-human and human-computer trust are different. One area relates to different cognitive schema's held by trustors, with schemas relating to automation technology being viewed as more perfect and schemas of humans being seen as fundamentally flawed. Next monitoring behaviors are said to differ between individuals monitoring automation with human monitors being more observant of errors committed in the automation condition than in the human condition. Last, the basis of trust in human-machine trust are said to be performance linked while the basis for trust in human to human contexts are said to be knowledge linked.

### 3. Theory and Research Questions

In human-robot interaction studies, increases in a robot's humanness have been correlated to increased perceptions of intelligence, comfort and even trust [10, 39]. More broadly, research in intelligent systems has shown anthropomorphism can preserve trust in the face of systems with deteriorating reliability [28]. Seymour et al. [25] found that humans rated a video human advisor as more trustworthy than either an avatar powered by AI or an avatar controlled by a human. No differences were observed in affinity, trustworthiness, or preference toward the avatar when comparing the one perceived to be controlled by AI or the avatar perceived to be controlled by a human. A different study looking at perceptions of human versus embodied computer teammates by Riedl et al. [23] showed that there were no behavioral differences between human and intelligent system behaviors, however there were significant differences in brain activation suggesting different cognitive processes involved in trust of each. It is unclear if differences in trust exist when interacting with a chat-based partner perceived as either human or an intelligent system. We therefore pose the following research question:

*RQ1: How does trust in a chat-based partner differ depending on the perception of humanness?*

Intelligent system partners may be perceived as human or not human either by directly informing users or through the conveyance (intentionally or unintentionally) of subtle cues over the course of an interaction. In regard to the former, research has been done exploring chat-based system identity disclosure from the onset of an interaction. Individuals experienced reduced perceptions of social presence and perceived humanness from systems that disclosed their machine identity [10]. With regard to the conveyance of cues that influence perceptions of humanness, research suggests that simple human-like visual cues are effective [8]. Similarly message interactivity, or the ability of a system to carry out interdependent message exchanges referencing prior information greatly influences perception of humanness [20]. In the absence of visual cues or in work where the nature of the task prevents high degrees of message interactivity, other cues are likely to influence an individual perception of humanness. In an effort to identify such cues, we propose the following research question:

*RQ2: What cues influence an individual's perception of a chat-based partner as human or not human?*

## 4. Method

### 4.1 Sample

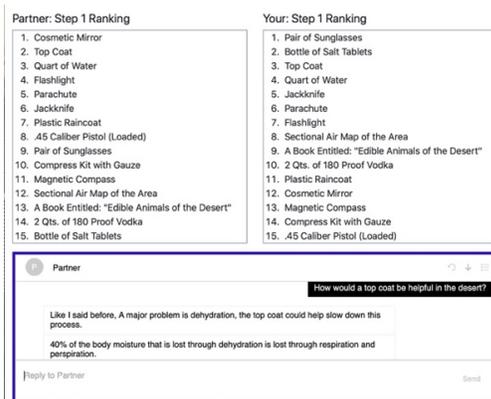
Participants were graduate and undergraduate students from a medium sized Midwestern university. A total of 48 subjects were recruited from a subject participant pool and compensated with course credit. Data collection occurred over a period of two months. Participants ages ranged from 19 to 24 years with the average age being 21 years.

### 4.2 Study task & apparatus

The tasks utilized in this study included the "Desert and Reef Survival Simulations" originally developed by Human Synergistics. These tasks were chosen because they had been previously utilized in numerous human subjects studies and had performance data for a number of populations. In addition, the specific survival situations involving desert and reef environments were specifically chosen

as they would be environments that were likely unfamiliar to participants from our sample population.

The “Desert and Reef Survival Situations” described scenarios where people had been stranded with only a limited number of items that could be used to survive. The goal of the simulations was to identify which of these items were most essential and rank the items in order of their importance for survival. For each survival simulation round participants would make two rankings, an individual ranking and then a final ranking that was made with consideration of solutions and input from a partner. After generating a ranking solution individually, participants were allowed to view their partner’s solution and converse with their partner to better understand the reasoning behind the partner solution (refer to Figure 1). Participants were told that their final ranking would be compared against a solution developed by expert military survival trainers. Participants were also informed that they would need to rank 75% or more of their items correctly (as compared to the expert’s ranking) or they would not receive participation credit for the study (in reality all participants received credit for their participation).



**Figure 1. Study platform showing partner chat interface.**

Participants were told that their partner would be a human survival expert who would be completing the simulation simultaneously and was not privy to the solution key. In reality, the partner was a chatbot programmed to respond to the participant questions about items from the survival scenarios in a conversational form. The chatbot was capable of understanding what item was being inquired about and provide detailed information regarding the items use and purpose. Information about each of the items was taken from the explanations provided in the survival simulation solutions manual developed by real world survival experts. Additionally, the chatbot was context aware and capable of generating a refreshed response

if participants had previously asked about an item. The system was capable of all of this without human intervention and could engage with participants fully autonomously. We developed a series of custom responses to handle messages that were out of bounds and redirect participants to ask questions that were about the items and to stick to the rules of the simulation.

### 4.3 Procedure

Participants completed the survival simulations and accompanying surveys online. Participants first completed an IRB mandated informed consent and then read a brief introduction where participants were made to believe that they were helping to evaluate a web application designed to aid decision making. Participants next completed the desert simulation (round 1) and immediately after left the study room to take a post task assessment. Following this, they completed the reef simulation activity (round 2) and took a final post task assessment. After completing both survival simulations and all accompanying surveys, participants were debriefed and thanked for their participation.

### 4.4 Measures

This study assessed trust as well as perceptions of partner performance, process, and purpose. To do this, we utilized the following measures:

*Trust* was assessed using the Empirically Derived (ED) scale developed by Jian et al [12]. The 12-item instrument conceptualizes trust as being comprised of two factors (trust & distrust). The trust factors of the scale include confidence, security, integrity, dependability, reliability, trust and familiarity. The distrust factors include deceptiveness, underhandedness, suspiciousness, wariness, and harm. Example question items include: “I am wary of my partner” and “I am confident in my partner.”

Specific dimensions of trust (performance, process, and purpose) were also measured using Chien et al. (2015) verified Inter-Cultural Trust Scale (ICTS). The 9-item instrument utilized a 5-point Likert-type scale with anchors strongly agree to strongly disagree. Example items includes: “My partner improves my performance,” “my partner uses appropriate methods to reach decisions,” and “I can always rely on my partner to ensure my performance.”

After completing both survival simulation activities, participants completed a post study survey. The survey asked open ended questions pertaining to the identity of the partner. Respondents indicated whether they perceived their partner to be an

automated chatbot or a human survival expert. The results from these questions were then separated into common themes using a Thematic Analysis approach [1:6].

## 5. Results

In this section we present findings from both the study data and open-ended survey response questions.

### 5.1 Statistical analysis

There were 34 participants who perceived their partner as a chatbot and 14 participants who perceived their partner as a human. An independent-samples t-test was run to determine if there were differences in perceptions of partner trust, performance, process, and purpose between individuals who perceived their partner as a chatbot and those who perceived their partner as a human. There were no outliers in the data, as assessed by inspection of a boxplot. Data was normally distributed and there was homogeneity of variances across the variables.

**Table 1. Summary of Results**

Measure	Round 1 - PerBot	Round 1 - PerHuman	Round 2 - PerBot	Round 2 - PerHuman
Trust	3.60	3.95	3.60*	4.14*
Performance	3.87	4.24	3.83**	4.57**
Process	4.05	4.43	3.95**	4.60**
Purpose	3.82*	4.26*	3.78**	4.43**

Before presenting our findings in detail, we will briefly summarize our results here and in Table 1. In the first round only scores for the trust dimension purpose were statistically significant, with higher scores reported by participants favoring the partner perceived as a human. In the second round, statistically significant differences were reported across all the variables, with greater reported scores on by participants perceiving their partner as human. In the contexts of this experiment, it appears that individuals trusted a perceived human partner more than a perceived intelligent system partner over a period following initial experience. Next, we discuss our findings in greater detail. Reported data are mean  $\pm$  standard deviation, unless otherwise stated.

In the first round of the study, individuals reported greater average trust and its corresponding dimensions of performance and process at levels approaching significance. Individuals perceiving their partner as human reported higher trust scores ( $3.95 \pm .60$ ) than individuals who perceived their partner as chatbot ( $3.60 \pm .61$ ). This finding was not statistically different

( $p = .081$ ). Individuals perceiving their partner as human reported greater performance scores ( $4.24 \pm .55$ ) than individuals perceiving their partner as chatbot ( $3.87 \pm .74$ ) a non-statistically different finding ( $p = .104$ ). Individuals perceiving their partner as human reported greater process scores ( $4.43 \pm .55$ ) than individuals perceiving their partner as chatbot ( $4.05 \pm .61$ ), a non-statistically different finding ( $p = .051$ ).

Individuals perceiving their partner as human, had higher purpose scores ( $4.26 \pm .66$ ) than those perceiving their partner as a chatbot ( $3.82 \pm .68$ ), a statistically significant difference of 1.03 (95% CI, -0.87 to -0.01),  $t(46) = -2.69$ ,  $p = .047$ ,  $d = .66$ .

In the second round, statistically significant differences were found across all variables favoring the partner perceived as human. Trust scores were higher for partners perceived as human ( $4.14 \pm .59$ ) than for partners perceived as a chatbot ( $3.61 \pm .64$ ), a statistically significant difference of 0.53 (95% CI, -0.93 to -0.04),  $t(46) = -2.04$ ,  $p = .01$ ,  $d = .85$ .

Performance scores were higher for partners perceived as human ( $4.57 \pm .62$ ) than for partners perceived as a chatbot ( $3.83 \pm .68$ ), a statistically significant difference of 0.74 (95% CI, -0.93 to -0.04),  $t(46) = -3.49$ ,  $p = .001$ ,  $d = 1.14$ .

Process scores were higher for partners perceived as human ( $4.60 \pm .47$ ) than for partners perceived as a chatbot ( $3.95 \pm .61$ ), a statistically significant difference of 0.65 (95% CI, -1.01 to -0.27),  $t(46) = -3.51$ ,  $p = .001$ ,  $d = 1.19$ .

Purpose scores were higher for partners perceived as human ( $4.43 \pm .61$ ) than for partners perceived as a chatbot ( $3.78 \pm .64$ ), a statistically significant difference of 0.65 (95% CI, -1.04 to -0.24),  $t(46) = -3.24$ ,  $p = .002$ ,  $d = 1.04$ .

### 5.1 Thematic analysis

A thematic analysis was performed on the open-ended survey responses asking individuals to provide additional insight into perceptions of their partner. In total 33 individuals responded to these questions. The table below (Table 2) lists the themes identified and the focus of the question the responses correspond to.

**Table 2. Emergent Themes**

Question Focus	Theme Count	Theme Names
Partner Perceived as Chatbot	4	Syntactical Requirements, Response Time, Conversational Tone, Unrealistic Availability
Partner Perceived as Human	1	Knowledge
Chatbot Improvement Suggestions	2	Natural Conversation, Realistic Response Time

*“The answers were all pre-written and determined. They never spoke about anything that I did not directly ask about.”*

### 5.1.1 Partner as Chatbot

First, we asked participants who indicated they believed their partner was a chatbot to explain their reasoning behind this belief. These individuals were asked: “What made you think your partner was an automated chatbot?”

Four themes emerged from participant responses to this question. The first theme “syntactical requirements” related to respondent’s belief that their partner imposed strict response requirements on them. In total 11 individuals were identified with responses fitting this theme. The following participant responses are illustrative of this theme.

*“I could only speak in specific ways. I spoke like I was supposed to several times and it still did not understand most likely due to a wrong letter being capitalized.”*

*“The need to use specific words when asking questions is also a sign. A human can infer things and respond to more than one question at a time, but a chat box cannot.”*

The second theme, “response time” related to participants belief that their the partner’s response time was too fast. In total, 9 individuals were identified in this theme. Examples from respondents who believed the partner demonstrated unnatural typing speeds are provided below.

*“The super quick response, no one types that fast.”*

*“The speed of the responses were too quick to be a person.”*

The third emergent them was “conversational tone.” In total responses from 7 individuals comprised this theme. Participants in this theme believed their partner was an automated chatbot because the partner’s responses did not demonstrate natural conversational dialogue or responses from the partner seemed too scripted. The following are quotes illustrating responses included in this theme.

*“There’s no doubt a bot was on the other end recognizing key words and then inputting a predetermined response”*

A final theme was “unrealistic availability.” In total 4 different participants had responses that fit this theme. The nature of this theme related to the partner having an unrealistic time frame of availability. Participants were able to complete this study with the partner at times of the day that are not typically associated with human availability. The following are direct quotes displaying respondent’s perceptions.

*“It is unrealistic to believe that a survival expert could be matched up in real time.”*

*“It is unlikely a real person would be answering these types of questions at 7:30 on a Sunday night”*

### 5.1.2 Partner as Human

We also asked participants who indicated they believed their partner was a human to explain their reasoning behind this belief. These individuals were asked: “What made you think your partner was a military survival expert?”

One predominant theme emerged from these responses. The theme “expert knowledge” reflected respondent beliefs that their partner was a human survival expert because the partner demonstrated sufficient survival knowledge. In total 6 individuals had responses belonging to this theme. The following are quotes are examples of responses that fit this theme.

*“The knowledge behind their reasonings for the rankings of each item. There was a lot of insight as to why certain items were ranked higher than others.”*

*“My partner had lots of knowledge and experience when answering the questions which lead me to believe that my partner was a military survival expert.”*

### 5.1.3 Chatbot Partner Improvement Suggestions

Before finishing the survey, we asked participants to provide constructive feedback for their partner. For individuals who believed their partner was a chatbot, we asked for participants to share ideas for making their partner be perceived as more humanlike. Specifically, we asked “If we told you that your partner was in-fact a human military survival expert,

what feedback would you give your partner to help them not be perceived as an automated chatbot?”

“Natural conversation” was a major theme that emerged and related to participants recommendation that the partner make their response more natural and conversational. Participants in this theme had responses that indicated their partner communicated in ways not traditionally ascribed to humans. In total, 17 respondents were identified with this theme. The following illustrate responses belonging to this theme.

*“Answer questions a bit less mechanically; the knowledge provided was insightful and good, but came across as rote and a pre-canned response.”*

*“They should be more relaxed. And even though the information are very good, there should be something more human like jokes or mistakes”*

Another theme related to improving chatbot partner perceptions was “realistic response time.” This related to the theme from the prior question asking reasons individuals believed their partner to be a chatbot. Individuals with responses belonging to this theme related to the partner having unnaturally fast response times. A total of 5 respondents were identified for this theme.

*“don’t answer questions so fast (autoprogrammed)”*

*“don’t respond as quickly. It’s not that that’s a bad thing, but it did give the impression of an automated chatbot”*

#### **5.1.4 Human Partner Improvement Suggestions**

Finally, we asked participants who believed their partner was a human, how their partner could improve. Specifically, we asked, “Is there any other feedback you would like to give your partner to help them be perceived in a more positive light?” Three major themes emerged from responses to this question: personal information, personalized response, and further engagement.

The first theme, “personal information” related to requests for the partner to present biographical information in order to be perceived as an actual human survival expert. Participants expressed their interest in knowing more personal information about the partner to make the conversational experience more natural; the following quote is illustrative of this theme.

*“I felt as if I perceived them in a positive light without knowing who they exactly were, but I feel as*

*if I would feel more easy knowing a little more about my partner since I am trusting their knowledge to help me pass.”*

A second theme, “personalized response” related to suggestions that the partner refer to the participant by their first name. Respondents believe this makes the conversation more personable; the following quote demonstrates this.

*“They could address someone by their name, and tell them not to worry or stay calm”*

The third theme, “further engagement” related to responses that indicated participants wished the partner provided further engagement throughout their interaction. Respondents want the partner to be less rigid when responding. The following quote is illustrative of this theme.

*“Have more human-like qualities such as making a joke to make the partner seem more human-like and have the partner ask questions too.”*

## **6. Discussion**

The results of this present effort significantly advance the understanding of trust in ambiguously identified systems. We observed statistically different levels of trust between systems perceived as human and systems perceived as machine-like, in the direction favoring systems perceived as human. In the context of chat-based system interfaces, this study is (to the best of our knowledge) the first to provide empirical evidence that, individuals who perceive their partner as human have greater levels of initial trust than those who perceive their partner as machine-like. If corroborated by other studies, this finding could have significant implications for system designers. For example, chat-based intelligent systems could be intentionally positioned or designed to promote perceived humanness in situations where higher levels of trust are desired. Also, our results showed statistically significant differences in trust only after the second period of performance with the partner. It is likely that over the course of the interaction, information cues either solidified the perception of working with a human partner or suggested the partner was non-human. Alternatively, it is possible that a third variable, such as an individual difference, drove the changes observed in either a partner’s perceived humanness or levels of trust. Future studies will want to directly manipulate perceived humanness to experimentally test the role of perceived humanness and trust.

Evidence of information cueing participants into their partners human/non-human identity were captured in the open-ended response to survey questions. Several important results emerged from the thematic analysis. First, individuals who perceived their partner as human reported that the reason for their belief was the perception of the system possessing expert knowledge. This is a significant contribution to the literature as it identifies a new strategy for system designers to promote perceptions of humanness above and beyond existing techniques like using human-like visual cues or increasing system message interactivity. Future studies should investigate this strategy further, looking to identify individuals who are most receptive to this technique and comparing its effectiveness to the other methods established for promoting perceptions of humanness.

In addition, our analysis identified several themes that support existing methods for manipulating perceptions of humanness in chat-based systems including the themes “conversational tone”, “response time”, and “syntactical requirements.” Another important finding was the theme relating to unrealistic availability. This finding can inform future research in ambiguously identified systems by restricting participation times to those that are reasonable for either human or a machine.

The study we have presented is not without its own set of limitations. First the system utilized in this study was intentionally restricted to respond to only relevant simulation-based questions. The partner therefore did not converse about topics such as their background, the weather, or other subjects outside the scope of the survival simulation. We recognize this limitation and feel it was necessary as this study is only the first of many iterations where the partner will be positioned in alternative ways with varying degrees of capabilities. Second, this study positioned the partner as human at the onset which may impact trust if individuals felt deceived or lied to. Future studies could iterate on this initial design in a number of ways, including manipulation of partner identity salience. Next, expertise of the partner was not manipulated, future studies could look at how this impacts perceived humanness. Finally, this study was conducted in a laboratory setting and like many empirical studies, results applied to other situations should keep this study’s context in mind.

## 7. Conclusion

Our findings suggest that, under the conditions of this study, chat-based intelligent system partners perceived as human are trusted at higher levels than those perceived to be machine-like.

## 8. References

- [1] Braun, V., and V. Clarke, “Using thematic analysis in psychology”, *Qualitative Research in Psychology* 3(2), 2006, pp. 77–101.
- [2] Byrne, D.E., *The attraction paradigm.*, Academic Press, New York, 1971.
- [3] Costa, A.C., C.A. Fulmer, and N.R. Anderson, “Trust in work teams: An integrative review, multilevel model, and future directions”, *Journal of Organizational Behavior* 39(2), 2018, pp. 169–184.
- [4] Deutsch, M., “Cooperation and trust: Some theoretical notes”, In *Nebraska Symposium on Motivation, 1962*. Univer. Nebraska Press, Oxford, England, 1962, 275–320.
- [5] Elson, J.S., “Toward a Theory of Early Trust in Intelligent Systems: Exploring Psychological Factors and Cognitive Processes”, 2019.  
<https://search.proquest.com/openview/45b57e8c21ba72564c3ea1cc07fb1770/1?pq-origsite=gscholar&cbl=18750&diss=y>
- [6] Elson, J.S., D. Derrick, and G. Ligon, “Examining trust and reliance in collaborations between humans and automated agents”, *Proceedings of the 51st Hawaii International Conference on System Sciences*, (2018).
- [7] Friedman, B., P.H. Khan, and D.C. Howe, “Trust online”, *Communications of the ACM* 43(12), 2000, pp. 34–40.
- [8] Go, E., and S.S. Sundar, “Humanizing chatbots: The effects of visual, identity and conversational cues on humanness perceptions”, *Computers in Human Behavior* 97, 2019, pp. 304–316.
- [9] Golembiewski, R.T., and M. McConkie, “The Centrality of Interpersonal Trust in Group Processes”, In *Theories of group processes*. 1975, 131–185.
- [10] Hendriks, F., C.X.J. Ou, A. Khodabandeh Amiri, and S. Bockting, “The Power of Computer-Mediated Communication Theories in Explaining the Effect of Chatbot Introduction on User Experience”, (2020).
- [11] Jarvenpaa, S.L., K. Knoll, and D.E. Leidner, “Is anybody out there? Antecedents of trust in global virtual teams”, *Journal of management information systems* 14(4), 1998, pp. 29–64.
- [12] Lee, J., and N. Moray, “Trust, control strategies and allocation of function in human-machine systems”, *Ergonomics* 35(10), 1992, pp. 1243–1270.
- [13] Lee, J.D., and K.A. See, “Trust in Automation: Designing for Appropriate Reliance”, *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46(1), 2004, pp. 50–80.
- [14] Lewicki, R.J., and B.B. Bunker, “Developing and maintaining trust in work relationships”, *Trust in organizations: Frontiers of theory and research* 114, 1996, pp. 139.
- [15] Luhmann, N., *Trust and Power*, John Wiley & Sons Inc, Ann Arbor, Mich, 1982.
- [16] Madhavan, P., and D.A. Wiegmann, “Similarities and differences between human–human and human–automation trust: an integrative review”, *Theoretical Issues in Ergonomics Science* 8(4), 2007, pp. 277–301.

- [17] Mayer, R.C., J.H. Davis, and F.D. Schoorman, "An Integrative Model of Organizational Trust", *The Academy of Management Review* 20(3), 1995, pp. 709.
- [18] Mcknight, D.H., M. Carter, J.B. Thatcher, and P.F. Clay, "Trust in a specific technology: An investigation of its components and measures", *ACM Transactions on Management Information Systems* 2(2), 2011, pp. 1–25.
- [19] McKnight, D.H., L.L. Cummings, and N.L. Chervany, "Initial trust formation in new organizational relationships", *Academy of Management review* 23(3), 1998, pp. 473–490.
- [20] Nabi, R.L., and M.B. Oliver, eds., *The SAGE Handbook of Media Processes and Effects*, SAGE Publications, Inc, Los Angeles, 2009.
- [21] Nass, C., J. Steuer, and E.R. Tauber, "Computers are social actors", *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1994), 72–78.
- [22] Norris, S.L., and R.L. Zweigenhaft, "Self-monitoring, trust, and commitment in romantic relationships", *The Journal of Social Psychology* 139(2), 1999, pp. 215–220.
- [23] Riedl, R., P. Mohr, P. Kenning, F. Davis, and H. Heekeren, "Trusting humans and avatars: Behavioral and neural evidence", 2011.
- [24] Rousseau, D.M., S.B. Sitkin, R.S. Burt, and C. Camerer, "Not So Different After All: A Cross-Discipline View Of Trust", *Academy of Management Review* 23(3), 1998, pp. 393–404.
- [25] Seymour, M., L. Yuan, A. Dennis, and K. Riemer, "Facing the Artificial: Understanding Affinity, Trustworthiness, and Preference for More Realistic Digital Humans", (2020).
- [26] Sheridan, T.B., and R.T. Hennessy, *Research and modeling of supervisory control behavior. Report of a workshop*, DTIC Document, 1984.
- [27] Turner, J.C., M.A. Hogg, P.J. Oakes, S.D. Reicher, and M.S. Wetherell, *Rediscovering the social group: A self-categorization theory.*, Basil Blackwell, 1987.
- [28] de Visser, E.J., S.S. Monfort, R. McKendrick, et al., "Almost human: Anthropomorphism increases trust resilience in cognitive agents.", *Journal of Experimental Psychology: Applied* 22(3), 2016, pp. 331–349.