Genie in the Bottle: Anthropomorphized Perceptions of Conversational Agents

Anastasia Kuzminykh[1], **Jenny Sun**[1], **Nivetha Govindaraju**[1], **Jeff Avery**[1] **and Edward Lank**[1,2]
[1]Cheriton School of Computer Science, University of Waterloo, Canada
[2]Inria & University of Lille, UMR 9189 - CRIStAL, France
<akuzminykh, jenny.sun.2, n2govind, jeffery.avery, lank>@uwaterloo.ca

ABSTRACT
This paper presents a qualitative multi-phase study seeking to identify patterns in users' anthropomorphized perceptions of conversational agents. Through a comparative analysis of behavioural perceptions and visual conceptions of three agents – Alexa, Google Assistant, and Siri – we first show that the perceptions of an agent's character are structured according to five categories: approachability, sentiment toward a user, professionalism, intelligence, and individuality. We then explore visualizations of the agents' appearance and discuss the specifics assigned to each agent. Finally, we analyze associative explanations for these perceptions. We demonstrate that the anthropomorphized behavioural and visual perceptions of agents yield structural consistency and discuss how these perceptions are linked with each other and system features.

Author Keywords
conversational agents, interaction, personification, behavioral, visual, anthropomorphism, user perception

CCS Concepts
•**Human-centered computing** → **Interaction design; HCI theory, concepts and models;** User studies;

INTRODUCTION
Due to a growing ecosystem of technologies that support conversational capabilities, more and more users are interacting with conversational agents on a daily basis [11], including such intelligent personal assistants as Siri, Google Assistant, and Amazon Alexa [9]. The devices supporting conversational agents, such as smartphones, tablets, and smart speakers, are often intended to be located in a particular environment (e.g. Google Home, Amazon Echo), and rely on users' voice interactions with a persona (an agent), who is deliberately designed to speak and act human-like. These characteristics integrate assistants in both technological and social spaces [49] of a user's life, e.g. by increasing perceptions of social presence [44], and encourage anthropomorphism – attributing lifelike qualities [18] – to conversational agents.

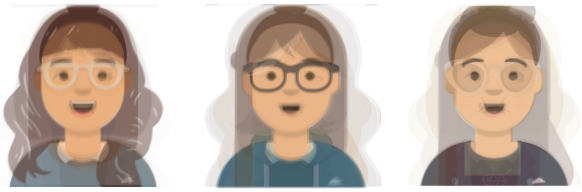

Figure 1. Cumulative visualizations. Left to right: Alexa, Google, Siri

While anthropomorphism has been shown to affect users' interactions, expectations, and overall satisfaction with conversational agents [42, 23, 26], it also introduces a number of unique adoption characteristics and integration challenges. For instance, there is evidence that the magnitude of anthropomorphization of agents and their social behavior shapes users' trust in the system, but anthropomorphic features can also increase trust resistance [13]. Furthermore, research suggests that users' satisfaction and relationship with intelligent agents are mediated by perceived trust and enjoyment during interactions [30], which are sensitive, among other factors, to the agent's interaction style [1]. Respectively, the interaction experiences with conversational agents might potentially be more important to users' satisfaction than the actual interaction output [33]. Thus, the understanding of processes that define the structure of anthropomorphized perceptions of conversational agents is essential for developing appropriate design of this technology and addressing their integration challenges. However, while research actively explores the degree of anthropomorphization of conversational agents in casual use [34, 23], and factors affecting the manifestation of anthropomorphism [29, 24, 42, 19], little is yet known on the structure and mechanisms of specifically anthropomorphized perception of these agents.

To address this gap, we focus on the structure and associative relationships within the anthropomorphized perceptions of conversational agents. Specifically, we examine the anthropomorphized perceptions of three popular agents to elicit patterns of perceptions through comparative analysis. Our results are drawn from a qualitative multi-phase study, leveraging semi-structured interviews primed with the interaction session with three agents, and followed by a visualization exercise. During the study sessions, participants were asked to compare specifics of each agent, and then explicitly asked to imagine and describe each of the agents as a human character.

CHI'20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
DOI: https://doi.org/10.1145/3313831.3376665

Research on anthropomorphic design, most developed for artificial agents and robotics [27, 28, 45], differentiates [14] between design of physical aspects of automatic agents (e.g. facial and body expressions [28, 51]) and design of their behavior traits. The behavioral aspects have been defined [25] as common and basic traits of human beings, including emotional responsiveness, interpersonal warmth, cognitive openness, individuality, and depth. By design, voice assistants are capable of displaying mainly behavioral anthropomorphic aspects, with the exception of voice, which can arguably be categorized as a physical trait. However, we hypothesized that behavioral anthropomorphized perceptions of conversational agents would potentially affect users' corresponding anthropomorphized visualizations, especially since an agent's answers might include visual references, e.g. "I'd blush if I could" [52]. Then, if there are patterns and consistent differences in behavioral perceptions of agents, we can also expect consistent differences in related visual conceptualizations of agents. To explore this hypothesis, we started by separately examining participants' descriptions of the anthropomorphized behavioral traits and the visual appearance of the agents.

Based on this analysis, we first demonstrate that the perceptions of an agent's behavior and character are structured according to five categories, where each category represents a comparative scale: the approachability category (emotionally/socially approachable versus distant), sentiment toward a user (genuine and caring versus disingenuous and cunning), professionalism (degree of manifestation), intelligence (positive or negative), and individuality (neutral versus defined personality). Respectively, we discuss the differences in the perceptions of three studied agents within these categories. We then present the results of visualizations of the agents' physical appearance, including body build and age, facial expressions, hair, and clothing style, describe the observed tendencies in the structure of descriptions in each of these categories, and discuss the consistent specifics assigned to each agent.

These results allow us to demonstrate the patterns appearing within both anthropomorphized behavioral and visual perceptions of conversational agents. To further validate and operationalize these findings, we then explore the associative relationships between different descriptive elements. Inspired by the concept of dynamic anthropomorphism [32] in robotics, we consider anthropomorphism of conversational agents as two-fold. The first level is represented by intended anthropomorphic design of an agent – explicitly built features, such as voice characteristics, richness of responses, or interaction style (e.g. jokes, laughter, answers to personal questions, etc.). The second level encompasses the actual process of anthropomorphization of an agent by a user, which emerges from the interaction. This refers to the anthropomorphized perceptions developed based on users' interpretations of designed features and further attribution of corresponding human characteristics, for example, being "patient", "genuine", "stressed", "motherly", "chill", etc. (examples from our data).

To gain insight into the relationship between anthropomorphized perceptions of agents behavior and visual appearance, and their association to system features, we analysed participants' self-reported explanations and associations between these types of elements. We discuss the specifics of these relationships, and demonstrate that the character perception is self-reportedly motivated only by the design, while visualizations are motivated by some design features, but are predominantly associated with the character's perceptions.

In the remainder of this paper, we first provide an overview of relevant research on anthropomorphism, followed by the multiphase study design and data analysis. After presenting our findings, we discuss them in the context of previous research and potential implications for conversational agent design.

RELATED WORK

Anthropomorphism, also referred to as personification, can be defined as people's tendency to attribute lifelike qualities, characteristics, or mental states to objects and other non-life artifacts [17, 21, 15, 50]. The media equation paradigm suggests that among other objects people specifically tend to apply social expectations to computers [40, 37], and "mindlessly" perceive computers as human-like actors, even when no explicit anthropomorphic features are involved [38].

However, a number of computational systems, including intelligent assistants, purposefully design anthropomorphic characteristics, particularly because anthropomorphism has been shown to ease user interactions. For example, Breazel et al. [6] showed that robots and other agents are more approachable and engaging when they exhibit human characteristics. Karjalainen et al. [26] studied the design space for drone companions and found that people were particularly positive towards a drone companion that featured anthropomorphic features. In a related vein, the emotion analysis of Amazon Echo reviews showed a link between personification of Echo and more positive emotions towards the agent, compared to those users who did not personify the device [24]. Earlier, Qiu and Benbasat [44] showed that using humanoid embodiment and human voice-based communication might enhance users' trust beliefs. Furthermore, Waytz et al. [50] demonstrated that individual differences in anthropomorphism predict the degree of moral care and concern afforded to an artificial agent, the amount of responsibility and trust placed on an agent, and the extent to which an agent serves as a source of social influence.

Despite the active development of the domain, the processes underlying the manifestation of anthropomorphization are still poorly understood. Some research suggests that anthropomorphized perceptions are binary and occur above a specific threshold, below which the agent is perceived as a computer [3, 13, 53]. Approaching this threshold can cause significant decreases in liking and trust around the technology displaying anthropomorphic features [36, 35]. On the other hand, there is evidence challenging this theory. In particular, Thompson et al. [48] found that ratings of the "humanness", familiarity, and eeriness of walking avatars changed continually, suggesting that users' cognitive response to anthropomorphic features of a system might develop linearly, as opposed to occurring after a particular threshold. Finally, the model of the dynamics of anthropomorphism in robotics [32] distinguishes between the anthropomorphic features designed within a system and anthropomorphism as a process of attributing lifelike qualities

as prompted by interaction with this system. Correspondingly, the model proposes to view anthropomorphization as a dynamic process following three stages: initialization, when anthropomorphization tends to increase; familiarization, associated with a decrease of anthropomorphic effects due to the user's acquired ability to predict the system's behavior; and stabilization, when anthropomorphic effects stabilize over a longer time. The measurements of a user's emotional response when interacting with conversational agents also suggest its dynamic nature during interaction stages [2].

While conversational agents include a number of explicit anthropomorphic features, findings on the actual degree of their casual anthropomorphization yield ambiguous results. For example, the analysis of user reviews of Amazon Echo showed that a significant number of reviewers were personifying Alexa as an assistant, a friend, or a family member [23]. Concurrently, another study explored the anthropomorphization of Alexa through a diary study, and found that less than half of study participants reported personification of the agent [34].

Nevertheless, research is actively revealing factors affecting the manifestation of anthropomorphization of intelligent agents. Among these factors, a humanized agent's name seems to be linked to higher personification [43], by allowing users to form more human-level relationships [24] with an agent, and to give them a social role [29]. Anthropomorphization might also occur to address the interaction contextual factors. In particular, it might help users to make sense of an unfamiliar situation when their understanding of a system is poorer than that of humans, to reduce feelings of insecurity when they seek to explain system behavior, or to establish lacking social connections [16, 17]. On the other hand, greater personification was demonstrated to co-occur with more social interactions with an agent. For instance, Purington et al. [43] demonstrated that people from multiple-member households are more likely to personify Amazon Echo than people who live alone. Research in robotics also suggests that anthropomorphization might be affected by the level of similarity between an artificial agent and a user. Specifically, Eyssel et al. [19] found that participants tend to feel closer and anthropomorphize more those robots who speak with the same gender voice as them.

Investigations of the specifics of anthropomorphized perceptions are, nevertheless, rather fragmentary. For instance, a number of studies have demonstrated that people tend to transfer social stereotypes, such as ethnicity [41] and gender biases [39, 4, 5, 20] onto anthropomorphized systems. It was also shown that the perception of an agent's personality and helpfulness has strong associations with the task performed with the system (carrying out procedures versus providing opinions), more so than with the agent's appearance [10]. Yet, little is known about the variations of anthropomorphized perceptions. Given a paucity of relevant data on anthropomorphized perceptions, this paper explores the structure and specifics of behavioral and visual conceptions of conversational agents, as well as related associative relationships with designed features.

STUDY DESIGN

Our results are drawn from a qualitative multi-phase exploratory study seeking to identify potential patterns in users' anthropomorphized perceptions and conceptions of conversational agents. The study probes these perceptions by leveraging semi-structured interviews, primed with an interaction session and followed by a visualization exercise. Each study session was audio recorded with the participant's consent.

Participants

In total, our sample included 20 participants (age 20-27, av.age. 23), 10 female and 10 male, with diverse levels of familiarity with the targeted agents. The rationale for the balanced self-reported gender representation in our sample is motivated by previous research, which suggests that users might feel psychologically closer to agents of the same gender as them [19], and thus we wanted to balance potential impact of participants' gender on our results. Furthermore, there is evidence that perceptions and interactions with conversational agents might be influenced by gender stereotypes [4, 5, 20]. While some participants did not own a device with an agent (Alexa n=17; Google Assistant n=10; Siri n=8), others owned such devices, but had never used the agent (Alexa n=1; Google Assistant n=3; Siri n=5). Two participants (P15, P20) did not own any of the three devices, while others owned at least one. Finally, the majority of those participants who regularly use Alexa (n=2), Google Assistant (n=7), or Siri (n=7) use the default voice setting (female, American). As an exception, P3 (male) uses a British female voice setting for Siri, and P1 (male) and P19 (female) use an American male voice setting for Siri. Our rationale for recruiting participants with different levels of familiarity with each agent is two-fold: we wanted to capture the differences in the anthropomorphized perceptions rooted in experienced versus novice in-situ interactions, as well as see the effects of direct in-situ comparisons of agents on existing preconceptions. To address potential inequalities in participants' experiences with specific agents, we primed each interview and visualization exercise with a semi-guided interaction session, as described in the study process section.

Study Process

We started each one-on-one session by asking participants to complete a questionnaire to collect their demographic data and capture their usage patterns with each conversational agent.

We then invited participants to engage in a semi-guided interaction session with conversational agents, using an Amazon Echo Dot (3rd Gen) Smart Speaker equipped with Alexa, a Google Home Mini smart home speaker powered by Google Assistant, and an Apple iPhone XS powered by Siri. Participants were provided with three groups of three conversational prompts (access to knowledge, intellect, and personality groups), and were instructed to choose and execute one prompt per group for each conversational agent. The prompts were motivated by the results from Sciuto et al. [46], showing that users tested Alexa's capabilities by asking questions from three categories: access to knowledge – factual questions, intellect – sophistication of accessible commands, and personality – agent's responses to non-factual prompts. The methodological benefits of the interaction session in-situ comparisons of three agents are such that later we were able to tease out differences as they were seen by participants. Furthermore, the mechanisms of forming a reliable perception of an agent is an

unexplored research question; however, there is evidence that it depends on performed tasks (e.g. [10]). This motivated us to structure the interaction sessions based on previously reported types of questions used to test an agent's capabilities [46].

The interaction session was followed by a semi-structured interview. Researchers would first prompt a participant to discuss their experience with each agent, and to describe and compare aspects, specific to each agent. Participants were then asked to imagine each of the agents as a human character and to describe what their definitive characteristics would be. Consequently, and priming the visualization exercise, we asked participants to describe how each agent would look like as a human character. During the interviews, participants would often re-initiate interactions, prompting several agents with the same command to better inform their descriptions.

In the next phase – a visualization exercise – participants were invited to create an approximate visual representation of Alexa, Google Assistant, and Siri using an open source web avatar generator[1]. We acknowledge that using predefined software for creating visualizations might introduce a number of potential restrictions to the visualization nuances. However, we chose to use a standardized platform to overcome participants' potential resistance to a hand drawing exercise, as well as to allow better comparison of the visualization outcomes between participants, regardless of their drawing skills. We also conducted the visualization exercise after interviewing participants about agents' appearance to ensure that limitations of the software would not affect their initial perceptions. Finally, we deliberately chose an avatar generator that had a broad and diverse representation of appearances.

We concluded with a debriefing interview to capture participants' general notes and comments on their study experience.

Data Preparation and Analysis

After fully transcribing audio recordings of each session (interaction sessions, interviews, visualization exercises, and debrief), we composed three data corpora for further analysis.

The first corpus included descriptions of agents' behavior and character, e.g. *"she has an authoritative vibe, but still friendly"* [P2], and originally consisted of 155 quotes. However, most of these quotes included a host of descriptive elements. For example, the quote *"Google is more genuine, and Siri is more transactional. Google is like more trusting. If it was a social setting, I would rather hang out with Google instead of Siri"* [P1] includes the following description elements – "genuine", "transactional", "trusting", "would hang out with", "would rather not hang out with". Thus, next, we extracted all the specific descriptive elements for each agent for the analysis. While some elements consisted of one-two words (as above), others were left longer (e.g. *"It's jarring when she makes a joke because it seems like she's pretending"* [P4]) to preserve semantic wholeness for thematic analysis. This resulted in 284 descriptions (Google n=102; Alexa n=97; Siri n=85).

The second corpus included descriptions of agents' appearance (186 initial quotes), e.g. *"I picture her hair as down, black or dark brown"* [P11]. We split the descriptions in these quotes (n=254) into one of the following: clothing style, e.g. *"hippy, colourful dress, really bright, green-ish colour"* [P16] (n=81); hair style, e.g. *"she always curls it, you know those wavy ones"* [P7] (n=48); hair color, e.g. *"bleach blond"* [P5] (n=44); age, e.g. *"Alexa is a younger person and Siri is her mom or something"* [P4] (n=40); build, e.g. *"Siri would be the smallest, Google the tallest, Alexa in-between"* [P10] (n=21), and facial expressions, e.g. *"a less friendly expression"* [P8] (n=20).

Finally, the third corpus of data included quotes with associative relationships between descriptive elements (e.g. *"Just seems stressed – they've got a lot of stuff and services, so their hair is a bit dishevelled"* [P7]). We then identified associative pairs of descriptive elements for each quote in this data corpus (e.g. "stressed – have a lot of stuff and services", "hair is dishevelled – stressed") so that the first element in the pair corresponds to the result, and the second corresponds to the reason ("A because of B"). We split the resulting 167 associative pairs into either character (n=65) or visual appearance (n=102) descriptions based on the first element in the pair.

Note that quotes, combining descriptions of character, appearance, and/or descriptive association (e.g. *"taller because she is daunting"* [P8]), appeared in several data corpora.

After the preparation of data was completed, we coded [12] descriptive elements in each corpus. We first performed open coding to identify themes in the first data corpus across all three agents. We then applied the resulting coding scheme to data on each agent separately for validation and to allow comparison between agents. Each phase of coding was initially performed by authors independently, then codes were discussed and refactored in consultation with the research team. The second corpus of data was analysed in a similar manner. The third corpus of data included associative pairs of character or visual appearance descriptions, and thus, was analysed using the coding scheme from the first two data corpora.

The visualization exercise outcomes were composed into collages (Fig.3), and appearances of chosen visual elements (Tab.2) were calculated. We also composed a cumulative visualization for each agent (Fig.1) by overlaying avatars: avatars were split based on the hair style, and overlaid from the least to the most represented style, gradually giving those layers higher to lower opacity respectively. The results were analyzed in comparison to the second corpus of verbal descriptions data.

RESULTS

In this section, we present our findings on the structure and specifics of behavioral and visual anthropomorphized perceptions of agents, followed by the analysis of the associative relationships between these perceptions, designed features, and perceptions of the corporate entity that produced the agent.

Character and Personality

We started by exploring the descriptions of agents' character and behavioral specifics[2]. We first analyzed the whole corpus

[1] Avataaars, open source software, https://getavataaars.com/

[2] Note: The terminology presented is elicited from interview data. The associated views and opinions are those of the participants and do not necessarily reflect the opinions and vocabulary of the authors.

of data, combining descriptions of all three agents, and identified five major themes, each representing a comparative scale for an agent's character. These themes include the agent's social approachability, projected sentiment toward the user, professionalism, intelligence, and specifics of the agent's individuality. We then applied the same coding scheme to the behavioral descriptions of each agent separately (Alexa n=97; Google n=102; Siri n=85), to validate these themes and to explore the differences in the assessment of the agents (Fig.2).

Approachability

The theme of approachability reflects an assessment of an agent as either distant or approachable. The "distance" end of this scale represents a perceived emotional or social distance with the agent's persona. This includes such assessments, as, for example, an agent being *"less of my friend"* [P4]; *"snarky"* [P9]; and *"closed off"* [P6]. We also observed that this assessment would regularly be given in contrast to a user's expectation or willingness to engage in closer/warmer relationships. For example: *"I said 'I love you' and they didn't reciprocate"* [P2]; *"very one-sided relationship... like they didn't want to talk to me, like "here's your answer and now go away""* [P1]; *"not interested in answering back"* [P11]. Correspondingly, the "approachable" side of the scale reflect users' perceptions of an agent as emotionally warm, receptive, friendly, and willing to interact with a user. As an illustration, consider the following descriptions: *"welcoming"* [P11]; *"like I've made that familial connection with them"* [P12]; *"feels more like a friend"* [P13]; *"outgoing"* [P2]; and *"chill"* [P19].

Sentiment Toward a User

The second theme – an agent's perceived sentiment toward a user – is defined through the dichotomy of genuine vs. disingenuous agent's attitudes. While the approachability theme reflects the social and emotional closeness, the sentiment theme refers to the perceived trustworthiness in attitudes demonstrated by an agent and how focused they are on a user. Specifically, the "genuine" aspect includes descriptions of agents' actions and attitude as benevolent, caring, and altruistic, e.g. *"coming from a good place"* [P1]; *"more sincere"* [P8]; *"went above and beyond by asking if I want more"* [P3]. These descriptions also include the perceived level of care in agents' actions and motivations, e.g. an agent being *"motherly"* [P20]; *"patient"* [P17]; *"accommodating"* [P5]; *"thoughtful"* [P3]; or *"really trying to help you"* [P16]. Opposing these caring, benevolent perceptions, the "disingenuous" side of the sentiment scale reflects assessments of an agent as cunning, indifferent, or inconsiderate. For example, descriptions such as *"sly"* [P8]; *"ingenuine"* [P1]; *"don't try to figure out what I need, they think they know what I need"* [P3]; *"pretending"* [P4]; *"uses her sass as a defense mechanism"* [P7].

Professionalism

The theme of professionalism, first, includes descriptions of an agent's communication style as in work settings, e.g. *"textbook assistant, no-strings attached relationship"* [P8]; *"it's weird to me when she makes a joke, because I visualize her as professional"* [P4]; *"she gives a lot of details in whatever you try to ask her, so she's a little more mature"* [P20]. Consequently, perceptions of professionalism affect a projected

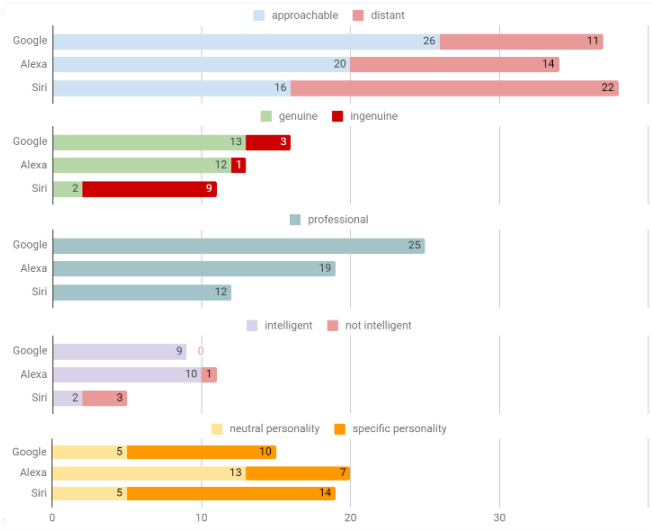

Figure 2. Number of character descriptions per agent in each category

type of relationship with the agent: *"Siri actually feels like an assistant. Like if you had someone following you around everywhere, doing favours for you, and giving you information that you need. And Alexa, just seems like a really smart friend that you have that can answer questions"* [P14]. This theme also includes perceptions regarding an agent's professional functioning, for example, such descriptions as *"self-sufficient"* [P14]; *"really organized"* [P4]; or *"like 'I gotta get stuff done'"* [P7]. The theme of professionalism differs from the previous two themes in that it does not include a defined negative value. However, while we have not encountered descriptions of specifically "unprofessional" behavior, we found that the comparisons between agents often reflect the differences in the degree of professionalism, especially in relation to the visualization of agents: *"Alexa has curly hair down, Google is like a bun, brown. Bun because more professional and strict"* [P10]; *"not as professional but still put together"* [P4]; *"isn't as put together, the other two are more sleek"* [P2].

Intelligence

The intelligence of an agent was mainly discussed in our data as an assessment of a positive value. For example, descriptions like *"witty with knowledge"* [P8]; *"well-read"* [P1]; or *"intelligent and sophisticated"* [P11]. It was significantly more rare for participants to describe an agent's negative intelligence (Fig.2), e.g. *"not intelligent"* [P5]; or *"a little on the dumber side"* [P20]. Due to the nature of the studied technology, which heavily focuses on information retrieval, the factor of intelligence might have be ambiguous notion in descriptions of an agent. In other words, a participant might be referring to the intelligence of an agent as a system or to the intelligence of an anthropomorphized persona of an agent. As an illustration, consider the following two descriptions of intelligence from the same participant: the system intelligence – *"sometimes says some weird stuff...you can tell it's trying to put sentences together, but it doesn't always make sense"* [P6] – and a persona intelligence – *"like the dumb kid that doesn't show up to class and asks everyone for help"* [P6]. In our analysis, we specifically focused on descriptions of anthropomorphized

intelligence, e.g. *"a little more wise, a little more knowledgeable"* [P20]; *"just sounds really smart, like someone really knowledgeable"* [P17], and disregarded commentary on the "system" intelligence when identifying the current theme.

Individuality

The last theme in the behavioral anthropomorphizations is an agent's individuality. This theme refers to participants' perceptions of how much character an agent is displaying and the specifics of this character. Respectively, this theme is represented by a dichotomy of neutral versus defined individuality, where specifics of defined individuality differ for each agent.

The neutral aspect includes descriptions that explicitly highlight the ordinariness of an agent's character: *"shy...not many interests"* [P7]; *"I feel like that's an average person"* [P9]; *"just boring... less inviting"* [P11]. This factor of ordinariness would often be associated with how noticeable the presence of the agent's persona is in the environment: *"It's just some random kid that would blend in"* [P6]; *"don't particular notice them more than the average"*; [P1] *"just wanted to blend in the background and not really be noticed"* [P14].

We then analysed the descriptions of defined individuality for each agent. We found that descriptions of Google Assistant's individuality revolve around the perception of a "nerdy" persona: *"really tall and lanky, because they're supposed to be like stereotypically nerdy"* [P7]; *"warmer but nerdy... nerdy chill"* [P6]; *"awkward... lanky. I'm just picturing someone who is a little bit plain but also a little awkward"* [P5]. Among three agents, Alexa's personality was most commonly described as neutral (Fig.2), "blending in the background" (e.g. *"normal, nothing too out there or different"* [P7]; *"she sounds kind of dull, like she's a dull worker behind a desk"* [P11]). The defined personality of Alexa was mainly described as peculiar and rather bright, e.g. *"she has a glowing personality"* [P16]; *"should be different from the other two. I feel like maybe... a bit more eccentric"* [P13]; and *"Alexa is more of a wild card"* [P8]). The descriptions of Siri's individuality are also portraying this agent as rather unusual with an accent on being audacious and daring, which is particularly noticeable in association with the agent's perceived appearance: *"she's hard to work with... I bet she's that way because she's really hot"* [P7]; *"[with] a bit of an edge"* [P4]; *"Siri is more hipster"* [P16].

Summary

Our analysis suggests that descriptions of an agent's character and aspects of behavioral anthropomorphism fall into one of five categories. Each category represents a comparative scale used to assess an agent: the approachability category (emotionally/socially approachable vs. distant), sentiment toward a user (genuine and caring vs. disingenuous, cunning attitude), professionalism (degree of manifestation), intelligence (positive vs. negative), and individuality (neutral vs. defined personality). Figure 2 illustrates the differences between agents' characters according to each of these five categories. In particular, Google's character tends to be described as approachable, closely followed by Alexa, but Siri was more commonly described as distant. Similarly, we observed a tendency to describe Google and Alexa as more genuine and caring, while Siri's sentiment was more commonly mentioned as disingenuous and inconsiderate. In our data, Siri's personality is commonly described as defined with particular individuality, while Alexa is noticeably more often seen as having a neutral, indistinct personality. Overall, we found it particularly interesting that the assessment of intelligence, commonly discussed in previous literature on conversational agents [47, 7, 8], is the least represented category of descriptions in our data.

Visualization of Agents

In the previous section we presented our results on anthropomorphized behavioral perceptions of conversational agents. In this section, we continue by discussing the visualizations of an agent's appearance. We first present the analysis of verbal descriptions of agents' physical traits[3], including physical build and age, facial expressions, hair, and clothes style (Tab.1). In particular, we describe the observed tendencies in the structure of descriptions in each of these categories, and discuss the specifics with regard to each agent. We then describe the analysis of the visualization exercise outcomes (Fig.1, Fig.3, and Tab.2) and present the consolidated results on the differences on visualizations of each conversational agent.

Physical Build and Age

In verbal descriptions of the physical traits of an agent's body, we found that some characteristics would yield rather consistent differences in perceptions of agents (e.g. height or age), while others would show no differences. Specifically, the overall build of agents would often be discussed as "average" and would not reveal any particular differences: *"fairly normal build, not super muscular... just very average"* [P1]; *"I see her as average North American built women"* [P8]; *"I picture them like thin, slim.... just average."* [P11]; *"I'm picturing a certain number of traits that are most... I don't want to say most common, but not different... very normal?"* [P13]

The factor of height, however, was a noticeable exception. In particular, participants would often note it as a comparative description, e.g. *"Same build, but Alexa maybe shorter"* [P11]; or *"Siri would be the smallest, Google would be the tallest, Alexa in-between."* [P10]. The height was most commonly associated with personality and behavioral perceptions of a specific agent, e.g.: *"Alexa slightly taller because daunting."* [P8], *"maybe Google seems taller. I don't know why, but that just seems more professional"* [P13]; or *"Alexa is shorter because she's less authoritative... and, I guess, I attribute that to tallness"* [P2]. Overall, we found a tendency for the Google Assistant persona to be generally perceived with a taller average height compared to the other two agents (Tab.1).

The descriptions of agents' age also revealed some differences in perceptions. Similar to height, age was often provided as a comparative description, e.g. *"Alexa maybe younger, out of all of them Siri is the oldest, but not by much, in 30s or mid 20s."* [P2]; *"Siri maybe later 20's, 30's, but Alexa 10-15 years older than that"* [P12]; *"Alexa is a little older, 30 to 40. Google is 25 to 30. I think, Siri is the youngest, 20-25."* [P20]. Correspondingly, age would often be associated with behavioral perceptions of an agent (e.g. *"she is younger, fresh*

[3]Since verbal descriptions were provided in a free form, not every aspect of visualization was mentioned by each participant.

Figure 3. Collage of visualizations of agents, organized by hair style. Left to right: Alexa, Google Assistant, Siri

	Element	*A*	*GA*	*S*		*Element*	*A*	*GA*	*S*		*Element*	*A*	*GA*	*S*		*Element*	*A*	*GA*	*S*
Gender	Male			1	Height	Short	**2**	1	1	Age	Younger	5	6	**9**	Face	Positive Express.	3	5	3
	Female	15	14	15		Average	3	**6**	**5**		Mid-Age	4	**8**	6		Neutral Express.	1	3	2
						Tall	1	**6**	1		Older	**8**	**8**	6		Negative Express.	1	2	1
Hair Style	Short	3		**5**	Hair Color	Silver	1	1	1	Clothes Style	Casual	**18**	10	10		Glasses	4	5	
	Wavy	**6**	2	3		Blond	1	**4**	**5**		Formal	7	**16**	**11**		No Glasses	1	1	1
	Straight	2	2	2		Brunette	**9**	**8**	2		Uniform-based	2	1	3					
	Updo	4	**15**	6		Dark	**3**	2	2		Company-based	1	**4**	1					
	Down	**10**	4	6		Black	**4**	1	**5**		Individuality	4	4	**10**					
	Other	1	1	**4**		Red	1												

Table 1. Frequency of visual appearance elements provided through verbal descriptions of Alexa (A), Google Assistant (GA), and Siri (S).

	Element & Type	*A*	*GA*	*S*		**Element & Type**	*A*	*GA*	*S*
Hair Style	Short	1	1	**5**	Hair Colour	Silver	1	1	2
	Long Straight	4	**7**	**7**		Blonde	0	**5**	**6**
	Long Wavy	**7**	3	2		Brunette	3	**5**	2
	Updo	4	**6**	3		Dark Brown	**10**	6	3
	Midlength	4	3	3		Black	5	3	**7**
						Red	1		
Gender	Female	19	19	16	Glasses	No glasses	13	13	16
	Male	1	1	4		Sunglasses	2	2	1
						Glasses	5	5	3
Eyes	Regular	12	13	12	Mouth	Open	10	9	10
	Happy	4	3	4		Open (teeth)	5	4	4
	Sad	1	1			Closed	2	1	1
	Other	1	1	2		Neutral	3	**6**	3
						Other			**2**
Cloth. Style	V/Scoop neck	**10**	4	**8**	Cloth. Colour	Dark/Black	11	10	12
	Hoodie		**4**			Blue	3	**5**	2
	Tank top			**2**		Grey	2	1	2
	Sweater w/collar	**6**	**5**	2		White/Beige	**4**	3	
	Jacket	**7**	**9**	**9**		Other		1	**4**

Table 2. Frequency of visual appearance elements provided through the visualization exercise for Alexa (A), Google Assistant (GA), and Siri (S).

grad vibes. Like trying too hard" [P6]). While we did not find any trends in a specific age group assigned to each particular agent, we observed a slight tendency for Siri to be perceived younger, and for Alexa to be perceived older (Tab.1): *"Alexa is older, in her late 20s. Siri and Google are in early 20s"* [P5]; *"Siri sounds younger, like late 20s. Alexa, I think, like 40s. Google sounds older... so between the two, mid 30s?"* [P9]; *"Alexa is the oldest, maybe late 30's to 40's... "* [P11].

Facial expressions

Perception of an agent's facial expressions seems, first, to be associated with how defined the agent's personality is for a particular participant, e.g. *"I guess, she looks deadpanned. If I could give her a face, I don't think I'd give her an expression"* [P4] versus *"She has an authoritative vibe... but still friendly, so still smiling, not stern"* [P2]. The specifics of facial expression often reflect the perceived approachability of an agent, for example, *"slight smile, because she is more playful"* [P3]. This is also reflected in the described differences in agents' facial expressions: *"Google just seems stressed, they've got a lot of stuff and services. Google is just focused on something, seems a little like their mind is somewhere else if you are having a conversation with them. Alexa is just maybe a little bit smiling, and Siri would be smiling a lot and has a lot of energy. In general, I can't imagine them being upset."* [P7]. Furthermore, participants commonly described agents' facial expressions based on their perceived role in an interaction: *"Alexa and Google have very neutral expressions until I prompt them, in which case it becomes a look of curiosity. In my mind, when they talk to me, it's a positive expression, and it can never be negative – I'm the client, they're doing me a service"* [P1].

Hair Style and Color

The hair styles verbally described for agents included different length, shape (*"looser curls"* [P4]; *"straight, medium/shorter length"* [P7]), style (*"tied up in a bun"* [P5]; *"down, long, wavy"* [P9]) and color (Tab.1). The descriptions of hair style appear to be significantly affected by perceived professionalism of an agent: *"really dressed-up person who's reporting something or like really formal, doing a really serious job. I associate that more, for females, having your hair up than down, at least for long hair."* [P17]; *"professional but also friendly, so hair in a bun or a ponytail"* [P2]; *"Alexa has curly hair down, Google has a bun, brown. Bun because more professional and strict"* [P10]. We have also encountered a number of very particular, definitive descriptions of hair, especially for Siri, e.g. *"weird short bangs"* [P15]; *"she's blonde but she always curls it. She got that Ariana Grande hair, half pony tail, really long"* [P7]; *"receding hairline"* [P6]. Such descriptions were commonly provided to highlight a perceived defined individuality of an agent. Similarly, the lack of individuality was also

reflected in hair descriptions: *"hair down, long, wavy? I feel like she is an average person and this is average"* [P9]; *"a normal ponytail, I just feel like it's very normal"* [P15].

Clothes and Overall Style

We categorized the descriptions of agents' appearance style as formal, casual, uniform-based, device/company inspired, and specifically reflecting agent's personality (Tab.1). For example, the following description includes a formal, a personality highlighting, and a device/company inspired perception: *"Siri, I would say, formal clothes, a work outfit, like a black suit, black pants, something like that. Alexa – jeans and sneakers and a t-shirt with an effect, a sentence. Something like "females are the future". And Google would be wearing like a Google t-shirt, jeans, and sneakers"* [P19]. Similar to other appearance elements, clothing style is especially descriptive when given as a comparison between agents: *"Google would wear a straw hat, wearing like a beach outfit. I picture a summer setting, because Google is in Mountain View. I feel like Siri is more hipster, less average. Siri wears like trendy overalls. And Alexa just wears jeans and a t-shirt"* [P15].

The uniform-based category includes descriptions of clothing style based on associations with a specific profession, for example, *"I associate like flight attendant clothes"* [P9]; *"Like when you're watching TV and you turn on the news, and there's someone sitting at the anchor. I guess, someone you'd expect to be reporting... like a journalist"* [P17]; *"like a news announcer... meteorologist on the Weather network"* [P14].

The category of personality highlighting clothing includes descriptions for each agent that mainly correspond to the themes of defined individuality of each agent (sec.4.1.5). For Siri, such clothing descriptions revolve around daring personality with accentuated femininity: *"Siri would wear heels, skirt, and a crop top, like a mini skirt."* [P7]; *"I almost picture Siri in like a school-girl outfit. Like a pleated skirt... like any teenage romcom"* [P5]; *"Some really hippy, colourful dress. And a really bright dress, green-ish colour."* [P16]. Google's clothes in this category are associated mainly with the perception of a stereotypically "awkward" persona: *"I imagine to have glasses. They could be wearing jeans and t-shirt but it doesn't fit them well"* [P7]; *"Google would dress like... glasses and the weird skirts and leotards. Sweater things, sweater vest things?"* [P20]. However, Alexa's clothes in this category do not seem to be consistent. As an illustration, compare the following descriptions: *"not regular human clothes because it doesn't feel human. Like sort of metallic clothes... maybe those reflective futuristic clothes"* [P13] and *"She's wearing mom jeans. And running shoes, because that's a very mom thing. Yeah, and glasses, that's a motherly thing too"* [P20].

Consolidation of Visualization Outcomes

We then compared the results of our analysis of verbal descriptions of agents' appearance with the outcomes of the visualization exercise. The visualizations of each agent were composed into a collage (Fig.3), and the number of appearance elements was calculated for each collage (Tab.2). We also created a cumulative visualization for each agent (Fig.1) to better illustrate the corresponding differences.

Our combined results suggest that Alexa tends to be perceived as being average height or slightly shorter, generally older than other agents, and wearing casual or business-casual clothes of dark or neutral colors. Her hair tends to be seen as darker, wavy, and worn down. Specifics of Alexa's appearance commonly highlight either a tendency to be unnoticeable (e.g. *"that's something I would wear if I just wanted to blend in the background and not really be noticed"* [P14]); cozy and home-associated visualization (e.g. *"a pullover sweater, skinny jeans, no shoes, I imagine her sitting on a couch"* [P4]); or an eccentric personality through futuristic clothes.

Google tends to be seen as being average height or taller, wearing either casual clothes with a focus on tech culture (e.g. hoodies), or business-formal clothes, both of dark or neutral colors. Google's hair tends to be seen of lighter color (blond, brunette) and as either long and straight, worn down or worn up (bun, ponytail), specifically associated with higher professionalism. The aspects of appearance highlighting their personality mainly revolve around the perception of a busy (*"she is not looking at me, she's busy"* [P4]), stressed (*"hair maybe gray? because of stress"* [P6]), or rather "awkward" person (*"a little bit plain but also a little awkward"* [P5]).

Siri is commonly described as being of an average height, younger than the other agents, and least often wearing glasses; wearing either casual but fashionable clothes (v-necks, tank tops, heels) or strictly business-formal style, of either dark or particularly bright colors, especially red. Siri's hair is described as short (most commonly among the three agents), or as long straight hair worn down, either blond or black, or with unique descriptions, e.g. *"black, but dyed black of white girls"* [P15]. Finally, descriptions of appearance highlighting specifics of Siri mainly portray a daring personality with accentuated femininity (e.g. *"the cheerleader trope"* [P7]).

Associations between Character and Appearance

Earlier, we discussed the structure and specifics of anthropomorphized perceptions of agents' character, as well as visualizations of their physical appearance. To further operationalize these results we now explore the associative relationships between different descriptive elements. For the analysis of these associations, we identified self-reported explanations (n=167) provided by our participants within their anthropomorphized descriptions (e.g. *"hair is down and wavy because she isn't as put together"* [P2]). In particular, we wanted to find the associations between behavioral and visual perceptions and the designed (objective and intended) features of each agent.

We first explored the designed features mentioned by participants as associative explanations, and identified four corresponding categories. First, (1) the category of provided information, including the amount of information and level of details (*"they give more info and so it seems like they care about you understanding what they're saying"* [P2]), as well as the very fact of providing the information (e.g. *"Siri is just cold, she copped out on a lot of answers and that made it more distant"* [P8]). Second, the designed (2) interaction style of an agent, such as responsiveness to prompts (e.g. *"I always have to be very clear with what I want from her, so she doesn't feel as flexible as a person"* [P1]), prompting users back (*"seems*

Design	*Approachab.* app.	*Approachab.* dist.	*Sentiment* gen.	*Sentiment* disin.	*Individ.* neut.	*Individ.* def.	*Prof.*	*Total*
Information	4	2	5		2			13
Interac. Style	11	10	5	6	1	1	4	38
Voice	3	5	1	1		1	1	12
Company	2							2
Total	*37*		*18*		*5*		*5*	

Table 3. Associative pairs for elements of character descriptions (columns), explained through designed features (rows).

Character		*Build*	*Cloth.*	*Face*	*Hair*	*Age*	*Total*
Approachability	approach.	1	6	2		1	16
	distant	2	2	1	1		
Individuality	neutral	2	7		6		33
	defined	2	4	2	10		
Professionalism	more	3	11	2	9		29
	less		1		3		
Sentiment	genuine		2				2
	Total	10	33	7	29	1	
Design		*Build*	*Cloth.*	*Face*	*Hair*	*Age*	*Total*
	company	2	6		1		9
	int. style		3			4	7
	voice		2			4	6
	Total	2	11		1	8	

Table 4. Associative pairs for elements of visual appearance (columns), explained through character and designed features (rows).

closed off, because doesn't do any prompts back like 'how are you' and stuff" [P6]), and the tendency to use jokes (*"when I asked what's her favourite colour, she told me a joke. So yes, way more playful"* [P3]). The third category of features refers to agent's (3) voice characteristics, e.g. *"More enthusiastic? like there's more inflection in her voice, sounds a bit more fun"* [P13]. Finally, we included the category of associations motivated by the (4) perception of the company, e.g. *"Alexa is more chill. Maybe because it's associated with a brand that reminds me of books and stuff"* [P19].

We then separately considered the associative explanations provided for elements of character (Tab.3) and visual appearance perceptions (Tab.4). We found that different categories of character perception are self-reportedly motivated only by designed features of agent technology (Tab.3), while visual appearance is motivated by both designed features and – predominantly – by character perceptions, with the exception of the agent's age. Note that the factor of intelligence did not appear in any of the elicited associative explanations.

In the character perception, the designed features are mainly affecting (Tab.3) the perception of agent's approachability, followed by the assessment of their sentiment toward the user. Both character categories are predominantly influenced by the interaction style of an agent, and the information they provide. Further, agent's approachability is associated with the agent's voice characteristics, e.g. *"less approachable, because the voice is just dull, doesn't seem interested in answering back"* [P11]. Overall, among the identified designed features, the agent's interaction style is most commonly associated with character perceptions, including – in order of extent of influence – agent's approachability (n=21), sentiment (n=11), professionalism (n=4), and individuality (n=2).

The visual appearance was predominantly motivated by the character perceptions (n=80). The most pronounced influence on visualization comes from the perception of the agent's individuality, both as defined (n=18) or neutral (n=15). The next most influential factor is the agent's perceived professionalism (n=29), followed by the agent's approachability, both when assessed as approachable (n=10) and distant (n=6). Interestingly, the perceived sentiment toward a user does not appear often in association with visual appearance, and intelligence of an agent does not appear at all. The most common aspects of appearance motivated by an agent's character include clothing style (n=33), followed by the agent's hair (n=29), and physical build (n=10). Perceived age of an agent, however, is mainly motivated by designed features, specifically by the interaction style and voice features of an agent. Designed features also seem to noticeably affect visualizations of an agent's clothing style, especially associations with the corresponding company. The design features mentioned as affecting visualizations include interaction style, voice, and the agent's parent company, while the information factor did not appear in our data.

DISCUSSION

Our findings suggest that the purposefully anthropomorphized behavioral and visualized perceptions of conversational agents yield structural consistency, discriminating differences between agents, and are linked with each other and system features though underlying associative relationships.

We found that the descriptions of behavioral anthropomorphism of conversational agents can be structured around five categories – approachability, sentiment, professionalism, intelligence, and individuality – each represented by a corresponding dichotomy. These results are particularly revealing when considered in the context of related theories. Exploring the concept of humanness and dehumanization, Harslam [25] describes two distinct senses of humanness: human uniqueness, believed to be acquired and to vary between people, and human nature, seen as fundamental characteristics "embedded" in the person, including cognitive flexibility, emotional responsiveness, individuality, depth, and interpersonal warmth. Harslam argues that the combination of attributed opposite characteristics (inertness, coldness, rigidity, fungibility, and lack of agency) represents a mechanistic dehumanization – a view of others as object-like or automaton-like. First, clear parallels between Harslam's human nature characteristics and some categories suggested in this paper (e.g. approachability – interpersonal warmth, sentiment – emotional responsiveness, individuality) validate the identified categories and the authenticity of the anthropomorphized nature of our data. On the other hand, the differences in the categories highlight the specifics of behavioral anthropomorphism of conversational agents. For instance, the behavioral category of professionalism suggests the tendency to project the role-based relationship onto the agent. While previous research also separates users' assessment of agents' access to knowledge and intellect [47], our results show that the anthropomorphized intelligence of

an agent is associated with knowledgeability, rather than cognitive flexibility, but the complexity of system performance affects either perceived professionalism or agents' sentiment.

The identified categories of behavioral descriptions also allowed us to compare the anthropomorphized perceptions of agents with respect to each category. For example, Siri's sentiment was predominantly described as disingenuous and cunning, while Alexa was described as genuine and caring. These particular results support findings from an earlier study [2] that assessed emotional responses to Alexa, Siri, and Cortana by measuring participants' biometric data, and concluded that Alexa generated the greatest emotional connection with participants. Contrarily to the sentiment perception, Alexa's individuality was commonly described as neutral and ordinary, while Google's and, especially, Siri's are considered to be more defined and pronounced. Finally, the variations in the relative "distance" between agents in each behavioral category both support the legitimacy of thematic clustering which defined these categories, and provide material for further exploration or the mechanisms underlying the differences.

Our results on the visualizations of purposefully anthropomorphized perceptions of agents demonstrate rather consistent differences in participant perceptions of agents' appearance, particularly in factors of their height, age, hair, and clothing style. We first would like to acknowledge that both character, and especially visual perceptions, elicited from the study, demonstrate an extreme tendency for displaying biases and social stereotypes. While this tendency is well established in anthropomorphized perceptions of computational systems [41, 4, 5, 20, 39], it might appear particularly defined for conversational agents due to their integration into users' personal and social environments [49]. Furthermore, a recent policy paper, produced by UNESCO [52], suggests that conversational agents potentially promote gender stereotypes by acting submissively in response to a user's displayed gender bias or even abuse. We have intentionally skirted analytical lenses in this paper, e.g. feminist theory, that might probe these aspects in more detail; this is an obvious avenue of future work.

In this paper, we rely on a comparative analysis, and present the specifics of each agent's visualization with the main purpose of demonstrating that visualization differences are present and consistent. Presenting these results, we mention behavioral connotations observed for particular forms of visual elements, e.g. hair worn up as a reflection of higher professionalism. However, due to the nature of the study, our data does not provide enough evidence to reliably establish detailed correlations. In essence, while our findings describe the structure of character and visual anthropomorphized perceptions, and demonstrate that they are interdependent, further research is required to confirm and expand on the understanding of specific dependencies between aspects of the agent's character perception and particular shape of visual elements. If established, these correlations could open exciting opportunities both for the system's feature design (e.g. designed answers referring to agent's visual appearance), as well as for the development of new projective investigation tools [31, 22] for research on perceptions of virtual agents.

As a step toward establishing these dependencies, we examined the self-reported associations between different categories (but not detailed forms) of anthropomorphized perception, and their relation to system features. Our results suggest that categories of character perception are only associated with the designed features of agents (information format, interaction style, voice characteristics, and agent's company), while visual appearance is motivated by some designed features (interaction style, voice characteristics, and agent's company), but mainly reflects perceptions of an agent's character. This both suggests the interdependence of agent character and visual conceptions, and the dominance of the behavioral anthropomorphized perception of conversational agents. These results also illustrate the differences of the relevant impact of system features on anthropomorphized perceptions, affecting mainly perceptions of agent approachability, sentiment, and age. Concurrently, visual perceptions are mainly associated with the agent's perceived individuality, professionalism, and approachability, and rarely with the agent's sentiment. Understanding these relationships between the identified categories of features, behavioral perceptions, and visualizations opens new avenues for a guided design of conversational agents.

LIMITATIONS

First, we acknowledge the effects of the diverse levels of participants' experience with conversational agents. Thus, while we found structural consistency in the descriptions of personality and visual appearance, we focus on reporting the specific attributes associated with a particular agent predominantly as an illustration of a theme. Correspondingly, our study design included the interaction session. While the main goal of the session was to equalize the recency of relevant stimuli for comparison rather than attempting to equalize the overall experience, the length of interaction sessions might also introduce some limitations in the results. Finally, we acknowledge that the benefits of standardizing the medium for visualizations for reliable comparisons can introduce limitations in the freedom of expression. To address the potential priming effect of the medium, we first conducted interviews, which, as reported, frequently showed *a priori* detailed perceptions of agents.

CONCLUSION

In this paper, we discussed the results of an exploratory study on the structure and interconnectedness of the anthropomorphized behavioural perceptions and visual conceptions of conversational agents. While, to the best of our knowledge, this study is the first attempt to identify patterns in the purposefully anthropomorphized perceptions of conversational agent technology, the consistent differences revealed through a comparative analysis of three agents suggest the importance of this research direction for the design of conversational agents.

ACKNOWLEDGEMENTS

We thank participants in our study. The research described in this paper was reviewed and approved by the Office of Research Ethics at the University of Waterloo. Funding for this research was provided by the Natural Science and Engineering Research Council of Canada (NSERC).

REFERENCES

[1] Timothy Bickmore and Justine Cassell. 2005. Social dialongue with embodied conversational agents. In *Advances in natural multimodal dialogue systems*. Springer, 23–54.

[2] Ben Bland. 2018. Alexa, Hug Me: Exploring Human-Machine Emotional Relations. (1 2018). Retrieved August 21, 2019 from https://medium.com/@ben.bland/alexa-hug-me-exploring-human-machine-emotional=-relations-1f0f6e04e1db.

[3] Jim Blascovich, Jack Loomis, Andrew C Beall, Kimberly R Swinth, Crystal L Hoyt, and Jeremy N Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry* 13, 2 (2002), 103–124.

[4] Sheryl Brahnam and Antonella De Angeli. 2012. Gender affordances of conversational agents. *Interacting with Computers* 24, 3 (2012), 139–153.

[5] Sheryl Brahnam and Margaret Weaver. 2015. Re/Framing Virtual Conversational Partners: A Feminist Critique and Tentative Move Towards a New Design Paradigm. In *International Conference of Design, User Experience, and Usability*. Springer, 172–183.

[6] Cynthia Breazeal. 2003. Toward sociable robots. *Robotics and Autonomous Systems* 42, 3 (2003), 167 – 175. DOI:http://dx.doi.org/https://doi.org/10.1016/S0921-8890(02)00373-1 Socially Interactive Robots.

[7] Raluca Budiu and Page Laubheimer. 2018a. Intelligent Assistants: Creepy, Childish, or a Tool? Users' Attitudes Toward Alexa, Google Assistant, and Siri. (8 2018). Retrieved February 22, 2019 from https://www.nngroup.com/articles/voice-assistant-attitudes/?lm=intelligent-assistants-poor-usability-high-adoption&pt=article.

[8] Raluca Budiu and Page Laubheimer. 2018b. Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri. (7 2018). Retrieved February 21, 2019 from https://www.nngroup.com/articles/intelligent-assistant-usability/.

[9] Raluca Budiu and Kathryn Whitenton. 2018. The Paradox of Intelligent Assistants: Poor Usability, High Adoption. (9 2018). Retrieved February 21, 2019 from https://www.nngroup.com/articles/intelligent-assistants-poor-usability-high-adoption/.

[10] Richard Catrambone, John Stasko, and Jun Xiao. 2002. Anthropomorphic agents as a user interface paradigm: Experimental findings and a framework for research. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 24.

[11] Pew Research Center. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones. (12 2017). Retrieved February 28, 2019 from https://www.pewresearch.org/fact-tank/2017/12/12/nearly-half-of-americans-use-digital-voice-assistants=-mostly-on-their-smartphones/.

[12] Kathy Charmaz. 2014. *Constructing grounded theory*. Sage.

[13] Ewart J de Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[14] Brian R Duffy. 2003a. Anthropomorphism and the social robot. *Robotics and autonomous systems* 42, 3-4 (2003), 177–190.

[15] Brian R Duffy. 2003b. Anthropomorphism and the social robot. *Robotics and autonomous systems* 42, 3-4 (2003), 177–190.

[16] Nicholas Epley, Scott Akalis, Adam Waytz, and John T Cacioppo. 2008. Creating social connection through inferential reproduction: Loneliness and perceived agency in gadgets, gods, and greyhounds. *Psychological science* 19, 2 (2008), 114–120.

[17] N. Epley, A. Waytz, and J. T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. In *Psychological Review*, Vol. 114. 864–886. DOI: http://dx.doi.org/10.1037/0033-295X.114.4.864

[18] F. Eyssel, L. de Ruiter, D. Kuchenbrandt, S. Bobinger, and F. Hegel. 2012. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 125–126. DOI: http://dx.doi.org/10.1145/2157689.2157717

[19] Friederike Eyssel, Laura De Ruiter, Dieta Kuchenbrandt, Simon Bobinger, and Frank Hegel. 2012. 'If you sound like me, you must be more human': On the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 125–126.

[20] Friederike Eyssel and Frank Hegel. 2012. (s) he's got the look: Gender stereotyping of robots 1. *Journal of Applied Social Psychology* 42, 9 (2012), 2213–2230.

[21] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In *International Conference on Social Robotics*. Springer, 199–208.

[22] Rosemary Flanagan and Robert W Motta. 2007. Figure drawings: A popular method. *Psychology in the Schools* 44, 3 (2007), 257–270.

[23] Yang Gao, Zhengyu Pan, Honghao Wang, and Guanling Chen. 2018a. Alexa, My Love: Analyzing Reviews of Amazon Echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, 372–380.

[24] Y. Gao, Z. Pan, H. Wang, and G. Chen. 2018b. Alexa, My Love: Analyzing Reviews of Amazon Echo. In *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. 372–380. DOI: http://dx.doi.org/10.1109/SmartWorld.2018.00094

[25] Nick Haslam. 2006. Dehumanization: An integrative review. *Personality and social psychology review* 10, 3 (2006), 252–264.

[26] Kari Daniel Karjalainen, Anna Elisabeth Sofia Romell, Photchara Ratsamee, Asim Evren Yantac, Morten Fjeld, and Mohammad Obaid. 2017. Social drone companion for the home environment: A user-centric exploration. In *Proceedings of the 5th International Conference on Human Agent Interaction*. ACM, 89–96.

[27] Aelee Kim, Jooyun Han, Younbo Jung, and Kwanmin Lee. 2013. The effects of familiarity and robot gesture on user acceptance of information. In *Human-Robot Interaction (HRI), 2013 8th ACM/IEEE International Conference on*. IEEE, 159–160.

[28] Bomyeong Kim, Hyun Young Kim, and Jinwoo Kim. 2016. Getting home safely with drone. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, 117–120.

[29] Cornell Social Media Lab. 2017. What Makes Us Love Alexa? (1 2017). Retrieved February 28, 2019 from https://medium.com/social-media-stories/what-makes-us-love-alexa-dd86e61be936.

[30] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[31] Marvin Leibowitz. 2013. *Interpreting projective drawings: A self-psychological approach*. Routledge.

[32] Séverin Lemaignan, Julia Fink, and Pierre Dillenbourg. 2014. The dynamics of anthropomorphism in robotics. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 226–227.

[33] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2018. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* (2018), 0961000618759414.

[34] Irene Lopatovska and Harriet Williams. 2018. Personification of the Amazon Alexa: BFF or a Mindless Companion. In *Proceedings of the 2018 Conference on Human Information Interaction&Retrieval*. ACM, 265–268.

[35] Karl F MacDorman and Debaleena Chattopadhyay. 2016. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* 146 (2016), 190–205.

[36] Karl F MacDorman and Hiroshi Ishiguro. 2006. The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies* 7, 3 (2006), 297–337.

[37] Clifford Nass and Kwan Min Lee. 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied* 7, 3 (2001), 171.

[38] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. 56, 1 (2000). DOI: http://dx.doi.org/10.1111/0022-4537.00153

[39] Clifford Nass, Youngme Moon, and Nancy Green. 1997. Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of applied social psychology* 27, 10 (1997), 864–876.

[40] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 72–78.

[41] Jean A. Pratt, Karina Hauser, Zsolt Ugray, and Olga Patterson. 2007. Looking at Human-computer Interface Design: Effects of Ethnicity in Computer Agents. *Interact. Comput.* 19, 4 (July 2007), 512–523. DOI: http://dx.doi.org/10.1016/j.intcom.2007.02.003

[42] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017a. "Alexa is My New BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2853–2859. DOI:http://dx.doi.org/10.1145/3027063.3053246

[43] Amanda Purington, Jessie G Taft, Shruti Sannon, Natalya N Bazarova, and Samuel Hardman Taylor. 2017b. Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 2853–2859.

[44] Lingyun Qiu and Izak Benbasat. 2009. Evaluating anthropomorphic product recommendation agents: A social relationship perspective to designing information systems. *Journal of Management Information Systems* 25, 4 (2009), 145–182.

[45] Laurel D Riek, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, and Peter Robinson. 2009. How anthropomorphism affects empathy toward robots. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. ACM, 245–246.

[46] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I. Hong. 2018a. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 857–868. DOI: http://dx.doi.org/10.1145/3196709.3196772

[47] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018b. Hey Alexa, What's Up?: A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 on Designing Interactive Systems Conference 2018*. ACM, 857–868.

[48] James C Thompson, J Gregory Trafton, and Patrick McKnight. 2011. The perception of humanness from the movements of synthetic agents. *Perception* 40, 6 (2011), 695–704.

[49] Alladi Venkatesh. 1996. Computers and Other Interactive Technologies for the Home. *Commun. ACM* 39, 12 (Dec. 1996), 47–54. DOI: http://dx.doi.org/10.1145/240483.240491

[50] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010a. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (2010), 219–232.

[51] Adam Waytz, Nicholas Epley, and John T Cacioppo. 2010b. Social cognition unbound: Insights into anthropomorphism and dehumanization. *Current Directions in Psychological Science* 19, 1 (2010), 58–62.

[52] Mark West, Rebecca Kraut, and Han Ei Chew. 2019. I'd blush if I could: closing gender divides in digital skills through education. (2019).

[53] Eva Wiese, Agnieszka Wykowska, Jan Zwickel, and Hermann J Müller. 2012. I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PloS one* 7, 9 (2012), e45391.