Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design

Isabel Kathleen Fornell Haugeland [a], Asbjørn Følstad [b,*], Cameron Taylor [c], Cato Alexander Bjørkli [a]

[a] *University of Oslo, Department of Psychology, Oslo, Norway*
[b] *SINTEF, Oslo, Norway*
[c] *Boost.ai, Sandnes, Norway*

ARTICLE INFO

ABSTRACT

Understanding the user experience of chatbots for customer service is essential to realize the potential of this technology. Such chatbots are typically designed for efficient and effective interactions, accentuating *pragmatic quality*, and there is a need to understand how to make these more pleasant and engaging, strengthening *hedonic quality*. One promising approach is to design for more humanlike chatbot interactions, that is, interactions resembling those of skilled customer service personnel. In a randomized experiment (n = 35) we investigated two chatbot interaction design features that may strengthen the impression of a humanlike character: (a) *topic-led conversations,* encouraging customer reflection, in contrast to *task-led conversations*, aiming for efficient goal completion, and (b) *free text interaction*, where users interact mainly using their own words, rather than *button interaction*, where users mainly interact through predefined answer alternatives. dependent variables were participant perceptions of *anthropomorphism* and *social presence*, two key concepts related to chatbot human likeness, in addition to *pragmatic quality* and *hedonic quality*. To further explore user perceptions of the interaction designs, the study also included semi-structured interviews. Topic-led conversations were found to strengthen anthropomorphism and hedonic quality. A similar effect was not found for free text interaction, reportedly due to lack in chatbot flexibility and adaptivity. Implications for theory and practice are suggested.

1. Introduction

Chatbots are machine agents that users interact with in natural language (Følstad and Brandtzaeg, 2020). Text-based chatbots are increasingly taken up for customer service (Følstad and Skjuve, 2019) as these represent an accessible and low threshold channel for users while being easy to implement and cost-efficient for companies (Shevat, 2017). A recent industry report (Gartner, 2019) found that 31% of customer communication managers have implemented chatbots or plan to do so in the near future and estimated that by 2025 customer service chatbots may increase operational efficiency by 25%.

Chatbots for customer service are typically designed with efficiency and effectiveness in mind (Nordheim et al., 2019). However, perceived quality in customer service may depend not only on frictionless goal completion but also on the emotional quality of the service experience (Berry et al., 2002). From a user experience perspective (Hassenzahl, 2018), chatbots for customer service are designed to maximize *pragmatic*

quality – that is, the character of the chatbot as useful and usable, serving the instrumental needs of the user. However, to realize the service quality potential of such chatbots, it may also be beneficial to strengthen *hedonic quality* – that is, the chatbot's ability to benefit user's well-being through engagement and stimulation. Strengthening hedonic quality in chatbots for customer service may, hence, strengthen user experience overall. Hedonic quality has been shown an important aspect of general chatbot user experience (Følstad and Brandtzaeg, 2020), and failures with chatbot applications have been attributed to a lack of engaging interactions (Jenkins et al., 2007; Schuetzler et al., 2014).

A promising approach for strengthening hedonic quality in chatbots, and thereby overall user experience, has been to strengthen their human likeness (Smestad and Volden, 2018). That is, to leverage chatbot features which make its interactions resemble those expected from a human (Araujo, 2018), utilizing human conversation as a metaphor for conversational design (Moore and Arar, 2018). For example, strengthening human likeness in chatbots for customer service could imply

* Corresponding author.
E-mail address: asf@sintef.no (A. Følstad).

designing chatbot interactions mimicking interactions between customers and skilled customer service personnel (Adam et al., 2020). Industry reports suggest that a substantial proportion of users expect humanlike characteristics, such as friendliness, in chatbots for customer service (Drift, 2018). Previous studies on trust in chatbots for customer service have found human likeness to likely be of importance for user experience (Nordheim et al., 2019). Human likeness in chatbots have been found to strengthen user perceptions of *anthropomorphism* and *social presence*; the former term referring to the chatbot being perceived as having humanlike traits (Araujo, 2018; Nass and Moon, 2000), the second referring to the chatbot being perceived as salient and immediate in its presentation and interactions (Go and Sundar, 2019).

Chatbot human likeness, and corresponding user perceptions of anthropomorphism and social presence, may depend on a range of interaction design decisions – from the chatbot persona and conversation style (Go and Sundar, 2019) to its conversational intelligence (Jain et al., 2018). Two aspects of interaction design currently understudied in this context are (a) the conversation types supported in the chatbot and (b) its interaction mechanisms.

A chatbot may support a number of *conversation types*, that is, forms of conversations with different style and objective. Roller et al., 2020b noted that chatbot conversations to communicate expertise and knowledge may span from goal-oriented task completion to in-depth discussions of specific topics. In the customer service context, conversations for goal-oriented task completion are critical (Xu et al., 2017). However, conversations that convey knowledge and information in an engaging manner are also desirable to users (Chung et al., 2020). Shevat (2017) captures this variation in conversation, distinguishing between *task-led* and *topic-led* conversations. The former concern narrow goal completion, the latter concern in-depth exploration and reflection on specific topics. Topic-led conversations, with exploratory and engaging exchanges between the user and chatbot, may contribute to a humanlike chatbot appearance as these more closely resemble informal human conversational interaction than do task-led conversations. As such, topic-led conversations may also add to the hedonic quality of chatbots for customer service.

The *interaction mechanisms* in a chatbot are the means with which users can send messages and receive information and content. These typically consist of a blend of free text input fields and buttons with predefined answer alternatives (Li et al., 2020; Shevat, 2017; Valério et al., 2017). Free text input may, arguably, enable interactions that strengthen the human likeness in the chatbot as these may resemble interactions with skilled customer service personnel. Button interaction on the other hand – though facilitating efficient interaction – may be seen as reducing users' perceptions of interacting with a humanlike entity (Jain et al., 2018; Valério et al., 2020). Hence, it is of high interest to know how increased use of free text interaction in customer service chatbots, without the support of button interaction, may strengthen its humanlike appearance and hedonic quality.

While conversation types and interaction mechanisms in chatbots for customer service arguably may impact human likeness and user experience, we lack empirical knowledge in this regard. To bridge this knowledge gap, the research objective for this study was as follows: To investigate whether and how manipulations of human likeness in chatbots for customer service – through variation in conversation types and interaction mechanisms – impact users' perceptions of the chatbot characteristics as well as its hedonic quality and, thereby, user experience.

To address this research objective, we conducted a 2×2 within-subjects factorial design experiment involving 35 participants. The factors were conversation type (*task-led* and *topic-led*) and interaction mechanism (*button interaction* and *free text interaction*). Dependent variables concerned user experience (*pragmatic quality* and *hedonic quality*) and the users' perceptions of the humanlike character of the interaction (*anthropomorphism* and *social presence*). To further investigate user perceptions of the different conditions, we also gathered data through semi-structured interviews.

The study findings help bridge the current knowledge gap regarding how user experience is affected by chatbot interaction design – specifically, how user experience is affected by changes in conversation types and interaction mechanisms assumed to strengthen the human likeness in the chatbot. As such, the study represents a needed contribution to the emerging area of human-chatbot interaction and holds implications for theory and practice.

The remainder of the paper is structured as follows. First, we present relevant background before explicating the research question and study hypotheses. We then detail the study method and present our findings, allowing the qualitative results to shed light on the outcomes of statistical hypothesis tests. Finally, we discuss the findings relative to the current state of the art, point out key implications, and consider limitations and avenues for future research.

2. Background

In the background section we present key characteristics and relevant knowledge concerning customer service and chatbots, before detailing relevant background on user experience research.

2.1. Customer service

Successful customer service is characterized by efficient and effective resolution of users' problems or requests (Dixon et al., 2010). However, efficiency and effectiveness may not be sufficient for an optimal user experience. Rather, companies aim to provide customer service that also generates positive emotions in users; seeking to please, engage and possibly surprise by going beyond users' expectations (Berry et al., 2002). User experience in customer service is affected by factors such as expectation management (Palmer, 2010), perceived courtesy, care, and friendliness (Hocutt et al., 2006), and adequately responding to customers conversation style (Liebrecht and van Hooijdonk, 2019). Hence, while customer service often has a practical outcome as its immediate objective, it may also aim for a positive emotional experience – a hedonic outcome.

The uptake of self-service technologies over the last few decades has fundamentally changed customer service delivery. Service theory in the eighties and nineties mainly discussed service as concerning the meeting between customers and skilled service personnel (e.g., Parasuraman et al., 1991). However, current service theory addresses multichannel service experiences and the impact of self-service technologies (Meuter et al., 2005).

Though customer service is increasingly conducted through digital self-service solutions, such as customer web pages and self-service smartphone apps, there is still a substantial need for access to skilled customer service personnel. This is evidenced by the marked growth in call-centres in recent years, in spite of the rapid and continuous improvement in solutions for self-service (Følstad et al., 2014). Customers often follow a multi-channel strategy when seeking help and information from companies, and studies suggest that about half of the customers that call customer support have first visited company websites (Dixon et al., 2010). In this context, chatbots for customer service may represent something in-between self-service on customer web pages and service involving skilled personnel. Interactions with chatbots for customer service may resemble conversational interactions with customer service personnel and may, hence, be perceived by users as more low-threshold and accessible than web page interactions (Følstad and Skjuve, 2019).

2.2. Chatbots

Computers that converse with users have been developed since the 1960's (Weizenbaum, 1966). We understand *chatbots* as 'software-based machine agents that provide access to services and information through

a conversational user interface' (Følstad and Brandtzaeg, 2020). While the term previously was used mainly in reference to agents for social chatter, it is increasingly used for task-oriented agents (e.g., Liao et al., 2018; Xu et al., 2017). Our use of the term is in line with how it is typically used in industry, though more inclusive than how the term typically is used in, for example, the natural language processing and dialogue systems communities (e.g., Jurafsky and Martin, 2021).

There has been substantial renewed interest in chatbots due to recent advances in artificial intelligence and natural language processing (Dale, 2016), as well as the uptake of chat platforms for personal, professional, and commercial communication (Følstad and Brandtzaeg, 2017). Chatbots are emerging in a range of application domains, such as health (Laranjo et al., 2018), education (Pérez et al., 2020), and work support (Meyer von Wolff et al., 2019). However, customer service remains one of the main application areas due to their promise to enable cost effective, available, and accessible service provision (Gartner, 2019; Kvale et al., 2020). Here, chatbots may be set up as a stand-alone support channel or as first line support, allowing for escalation to skilled customer service personnel when the chatbot cannot resolve the issues (Wilson and Daugherty, 2018).

2.2.1. Chatbot conversation types

A range of conversation types may be supported in chatbots. Whereas a coarse distinction has previously been made between chatbots for social chatter and dialogue systems for task completion (Chen et al., 2017), recent progress in open domain chatbots (Adiwardana et al., 2020; Roller et al., 2020a) suggests the benefit of supporting a range of conversation types in chatbots. In their outlook on future research directions, Roller et al., 2020b argue the need to move beyond narrow task-completion and recall of knowledge in response to frequently asked questions, towards also supporting knowledgeable in-depth conversations in chatbots.

Chatbots for customer service are typically designed for conversations with a narrow goal orientation (Forrester, 2016). Furthermore, studies of chatbots for customer service often concern chatbot characteristics in support of such task completion (Ashktorab et al., 2019; Kvale et al., 2020). While such task-oriented conversations may be seen as mechanical or scripted, they align with users' productivity motivations for using chatbots (Brandtzaeg and Følstad, 2017).

However, conversations corresponding to the in-depth knowledgeable conversations envisioned by Roller et al., 2020b are emerging also in the customer service domain. For example, such conversations have been explored to strengthen user engagement with brands (Chung et al., 2020) or to promote products, services, and informational content (Shevat, 2017).

To distinguish dialogues with a narrow goal orientation from dialogues for in-depth treatment of a topic, Shevat introduced the notion of task-led and topic-led conversations. The former conversation type is targeted and goal-oriented whereas the latter denote conversations exploring or detailing a topic of interest.

Conversation type may impact users' perceptions of chatbot human likeness. Topic-led conversations typically have high resemblance to interaction with human conversation partners, as is seen in their prevalence in conversations for relation formation (Ta et al., 2020) and engagement (Ram et al., 2018). As such, the topic-led conversations may be expected to strengthen hedonic quality through facilitating in-depth conversational interaction and an engaging conversation style resembling human conversationalists. Conversely, task-led conversations, where goal achievement is given priority – possibly at the cost of human likeness, will likely be perceived as high in pragmatic quality.

2.2.2. Chatbot input mechanisms

Chatbots may provide users with a variety of input mechanisms. While some chatbots rely exclusively on free text input from users (e.g., Adiwardana et al., 2020; Roller, et al., 2020), chatbots often also employ interaction mechanisms such as buttons, quick replies, and menus for efficient and effective interaction (Jain et al., 2018; Valério et al., 2017).

State of the practice chatbots for customer service typically allow users to enter their requests in free text and also provide buttons or quick replies to facilitate the dialogue (e.g., Kvale et al., 2020; Li et al., 2020). The free text input is processed for intent identification, based on machine learning models from text-based training data associated with each intent. When intent is identified, the interaction design is typically set up so as to efficiently funnel the user towards goal completion by way of a rule-based dialogue tree structure (Shevat, 2017). The alternatives at each node in the dialogue tree are typically presented as predefined quick replies or buttons, for efficient input of user information, reduced risk of erroneous interpretation of user input, and adaptations of chatbot content. In a study of chatbot interaction design and user experience, Jain et al. (2018) found users to appreciate the efficiency and ease involved in button interaction. However, chatbots relying exclusively on button interaction were in the same study found annoying to users, while natural language understanding in chatbots was found to be delightful. In a comparative study of chatbots for movie information (Valério et al., 2020), users considered buttons a valuable means for efficient interaction and also a means of signalling chatbot features.

Choice of interaction mechanism may impact chatbot human likeness. Arguably, free text interaction may provide users with a stronger sense of flexibility and engagement in the conversation, compared to button interaction. Such a sense of flexibility and engagement may suggest conversational intelligence in the chatbot thereby, in line with Jain et al. (2018), strengthen user perceptions of human likeness. Also, the use of free text interaction in open domain chatbots, intended to provide a conversational experience resembling that of a human conversationalist (e.g., Adiwardana et al., 2020; Roller, et al., 2020), may be indicative of the humanlike character of free text interaction. As such, it may seem reasonable that free text interaction will strengthen hedonic quality whereas button interaction may be associated with high pragmatic quality.

2.3. User experience

User experience concerns users' emotions, preferences, perceptions, and responses before, during or after the use of an interactive system (ISO 2019), and can refer to a broad range of phenomena (Law and van Schaik, 2010). Hassenzahl's (2018) hedonic-pragmatic model of user experience has been formative for much of the current research and practice in this field. Here, user experience is analysed as impacted by two main factors. *Pragmatic quality* concerns the usefulness, effectiveness, and efficiency of the system. *Hedonic quality* concerns the pleasurable aspects of interaction – including aspects such as the systems' capacity for stimulation, identification and evocation (Hassenzahl, 2018). The hedonic-pragmatic model is operationalized in the AttrakDiff instrument, where hedonic and pragmatic quality is measured through a set of validated questionnaire items (Hassenzahl et al., 2003). AttrakDiff has been applied to measure user experience in systems such as websites (Papachristos, 2019), smartphones (Kujala, and Miron-Shatz, 2013), games (Bernhaupt et al., 2007), and also chatbot prototypes (Smestad and Volden, 2018). Following from the hedonic-pragmatic model, solutions that have strong pragmatic and hedonic qualities are likely to be desired by users (Hassenzahl, 2018).

The hedonic-pragmatic model provides a perspective for understanding how user experience may be strengthened in chatbots, an idea further supported by Følstad and Brandtzaeg (2020). In a survey study of chatbot users, they found chatbots' pragmatic capabilities to significantly impact user experience. At the same time, capabilities to entertain, exhibit novelty, provide inspiration, and support social interaction were also found to strengthen user experience. Whereas pragmatic user experience is a key concern in current design of chatbots for customer service (Følstad and Skjuve, 2019; van der Goot et al., 2020), studies have found beneficial effects of aspects associated with hedonic quality.

Introducing humanlike characteristics in chatbots is a promising

means of strengthening hedonic quality and, thereby, user experience. Go and Sundar (2019) found that humanlike design cues in chatbots, specifically message interactivity, may impact users' perceptions of anthropomorphism, social presence and, thereby, perceived expertise and friendliness in the chatbot. Lee and Choi (2017) found that humanlike social behaviours in chatbots, such as self-disclosure and reciprocation, may increase enjoyment and user satisfaction. Schuetzler et al. (2018) found variations in chatbot conversational content to impact users' perceptions of anthropomorphism in the chatbot and, thereby, user engagement. Araujo (2018) found humanlike cues in chatbot interaction to strengthen users' perceptions of anthropomorphism in the chatbot and also their emotional connection with the company hosting the chatbot. However, Araujo did not find humanlike cues to strengthen social presence. Diederich et al. (2019) found a conversational agent adapting its responses according to the user sentiment to be perceived as more anthropomorphic, socially present, and to yield a higher satisfaction with the service encounter.

While human likeness in chatbot interaction design is expected to strengthen user perceptions of anthropomorphism, as well as social presence (Araujo, 2018), the current body of knowledge only addresses some types of design choices that may impact human likeness, such as chatbot identity, visual presentation, and message interactivity (e.g., Araujo, 2018; Go and Sundar, 2019). Hence, research is needed on how a wider range of humanlike chatbot characteristics – such as choice of interaction mechanism and conversation type – may affect perceptions of anthropomorphism and social presence, as well as hedonic quality and user experience.

3. Research question and hypotheses

In response to the need for knowledge on how human likeness in chatbot interaction design may affect user experience, we set out to systematically investigate two aspects of particular relevance for current chatbots for customer service – both which are currently understudied with regard to user experience. That is, (a) conversation types and (b) interaction mechanisms. The research question established for the study was:

How is the user experience of customer service chatbots impacted by variations in conversation type and interaction mechanism assumed to have implications for chatbot human likeness?

We were, in particular, interested in the impact that conversation types and interaction mechanisms may have on the two main aspects of user experience as defined in the Hassenzahl (2018) hedonic-pragmatic model. However, given our aim to study the effect of strengthening the humanlike character of the interaction mechanism and conversation type, we were also interested in the impact of these on users' perceptions of anthropomorphism and social presence of the chatbot.

3.1. Hypotheses for the impact of variations in conversation types

Choice of conversation type seems a likely candidate for impacting user experience. A chatbot with task-led conversations, aimed at efficient achievement of clearly defined goals, is likely to be perceived as high in pragmatic quality. However, such a chatbot may arguably be perceived as low in anthropomorphism and social presence given its processual character, and possibly also low in hedonic quality. In contrast, a chatbot with topic-led conversations, aiming for engaging exploration of a particular topic or area of interest, seems more likely to induce perceptions of anthropomorphism and socially presence, and possibly also increased hedonic quality.

In consequence, our hypotheses for the impact of conversation type are as follows:

H1. Topic-led conversations increase the perceived (a) anthropomorphism and (b) social presence in the chatbot compared to task-led conversations.

H2. Topic-led conversations increase the hedonic quality of the chatbot compared to task-led conversations.

H3. Task-led conversations increase the pragmatic quality of the chatbot compared to topic-led conversations.

3.2. Hypotheses for the impact of variations in interaction mechanisms

The choice of interaction mechanisms is also likely to impact users' perceptions of anthropomorphism and social presence in the chatbot as well as its pragmatic and hedonic quality. Button interaction may enable efficient chatbot dialogues and, hence, strengthen pragmatic quality (Jain et al., 2018; Valério et al., 2020). However, the efficiency and convenience of button interaction in chatbots may come at the cost of the chatbot being framed as a less humanlike conversational partner and potentially also reduce users' sense of engagement. Vice versa, free text interaction may strengthen the humanlike character in the chatbot and also hedonic quality. In consequence, our hypotheses for the impact of variations in interaction mechanisms are as follows:

H4. Free text interaction increases the perceived (a) anthropomorphism and (b) social presence in the chatbot compared to button interaction.

H5. Button interaction increases the pragmatic quality of the chatbot compared to free text interaction.

H6. Free text interaction increases the hedonic quality of the chatbot compared to button interaction.

4. Method

4.1. Research design

In response to our research question and hypotheses, we chose an experimental design set up as a within subjects 2 × 2 factorial design with randomization to conditions (Shadish et al., 2002). The factors were conversation type (task-led vs. topic-led) and interaction mechanism (button interaction vs. free text interaction). This choice of experimental design allowed for efficient investigation of both the studies factors at acceptable power even with relatively few participants. A within subjects t-test with 35 participants is expected to detect medium effect sizes with statistical power $1\text{-}\beta = 0{,}8$[1]. It should be noted that our goal was not to investigate any interaction effects between the factors; the factors were investigated independently.

The experimental design was augmented with qualitative exploratory data collection to gain further insight into the detailed user perceptions of the studied conditions of chatbot interaction design. This part of the study was implemented as semi-structured interviews, where the participants reported on their experience with the chatbot interactions.

The context of the experimental design was customer service in retail banking, one of the main sectors for uptake of chatbots during the recent years and a forerunner in digitalization of consumer services (Taylor et al., 2020).

---

[1] Power analysis conducted by the software G*Power 3.11 https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower.html

4.2. Recruitment, participants, and ethical considerations

Participants were recruited through an agency[2] and participated in person at the premises of the second author's research organization. In total, 35 participants were included, 17 males and 18 females. Participant age ranged from 20 to 60 years (mean = 39, SD = 12). All participants were bank customers and were either employed or students. During the recruitment process, potential invitees with a background in information technology or customer service were screened out. As incentives, participants received gift cards with an approximate value of 50 Euros.

Recruitment and participation followed relevant protocols of ethics and privacy in research, approved by the relevant data protection body. Participation followed informed consent, participants were informed they could terminate their participation at any time, and all data were treated confidentially and anonymized.

4.3. Chatbots, measurements, and interview guide

4.3.1. Chatbots

We set up four chatbot versions to instantiate the experimental conditions. The chatbots were based on the same platform and enabled either a task-led or a topic-led conversation and employed either button interaction or free text interaction.

The task-led conversation concerned a scenario where participants used a chatbot to order a credit card; a highly goal-oriented process. The topic-led conversation concerned a scenario where the participants used a chatbot for advice on pension savings; a conversation set up to encourage user reflection and engagement on the topic at hand.

For each scenario (task-led and topic-led), chatbots were implemented in two versions – one with button interaction and one with free text interaction.

Button-interaction was implemented so that all user interaction following the initial user request could be conducted mainly by selecting amongst buttons with predefined answer alternatives. Free text interaction was implemented so that the user had to provide all input in free text.

4.3.2. Questionnaire measurements

The dependant variables were perceived anthropomorphism, social presence, hedonic quality, and pragmatic quality. *Anthropomorphism* was measured with a scale adapted from Araujo (2018), consisting of three semantic differential items. *Social presence* was measured with a scale adapted from Laban and Araujo, 2020 consisting of four Likert scale items. *Hedonic quality* and *pragmatic quality* were measured by AttrakDiff, a verified scale for measuring user experience (Hassenzahl et al., 2003). The AttrakDiff scale was adapted by removing three items prior to data collection due to these not being found suitable for the studied context. For hedonic quality, we combined the scores for the AttrakDiff constructs *hedonic quality – identity* and *hedonic quality – stimulation*. The participants responded on the scales through dedicated questionnaire forms. The questionnaire items belonging to each measurement instrument was found to have substantial or near perfect agreement following Landis and Koch, (1977) rule of thumb, with values for Cohen's kappa ranging from 0.78 to 0.92. The questionnaire items for all measurement instruments are presented in Appendix A.

4.3.3. Interview guide

The participants were engaged in brief interim interviews following each chatbot interaction. For this purpose, the interview guide included two questions on their immediate experience of the chatbot and related chatbot use. Upon completion of both interactions and interim interviews, the participants engaged in a final interview on the two

chatbots they had tried in terms of similarities and differences as well as the factors of particular interest to the study; anthropomorphism, social presence, pragmatic quality, and hedonic quality. At the end of the final interview, the participants were specifically asked to reflect also on the use of button and free text interaction and their view of the two different scenarios (task-led and topic-led).

The research design was piloted with four persons prior to data collection. Based on the piloting, minor improvements were made to the chatbots and interview guide. An overview of the interview guide questions is provided in Appendix A.

4.4. Data collection and analysis

Each data collection session involved one participant and was led by a study moderator (the first author). Data collection included the following steps:

- interaction with first chatbot
- questionnaire scales for first chatbot
- interim interview, experience with first chatbot
- interaction with second chatbot
- questionnaire scales for second chatbot
- interim interview, experience with second chatbot
- final interview

All participants interacted with two of the four chatbots: one chatbot in the scenario for task-led conversation and one in the scenario for topic-led conversation. Of these, one chatbot would offer button interaction and one free text interaction. To counter any order effects, the participants tried the two chatbots assigned to them in one of four configurations: (a) Task-led conversation with buttons – topic-led conversation with free text, (b) task-led conversation with free text – topic-led conversation with buttons, (c) topic-led conversation with buttons – task-led conversation with free text, and (d) topic-led conversation with free text – task-led conversation with buttons. Participant 1, 5, 9 … used configuration (a), participant 2,6,10 … used configuration (b), etc.

The procedure for data collection lasted approximately 1,5 hours. All chatbot interaction data were stored on the chatbot server. All questionnaire responses were gathered and transferred to a common data sheet. All interviews were audio recorded and transcribed. All data were anonymized as part of the transcription and data preparation process.

Statistical hypothesis testing was conducted in the software package SPSS 26. The study hypotheses were tested by paired-samples t-tests. As outliers were observed for some of the variables, the analyses were repeated by the non-parametric Wilcoxon signed ranks test. The interview data were analysed through thematic analysis as described by Braun and Clarke (2006). The analysis process involved the first and second author. The first author established an initial set of themes and subthemes specific to the four chatbot conditions. The second author reviewed and merged these, resulting in a set of 32 subthemes grouped in 17 themes applicable across all four chatbot conditions. The final analysis and coding were conducted by the second author. Key themes are presented in the results section.

As some of the participants experienced usability issues when interacting with the chatbots, such as having to repeat their request or receiving false positive answers, we included additional analyses to check that such issues did not bias the results. For this, all interactions were rated for interaction problems. Then, all hypothesis tests were replicated on a data set excluding all participants experiencing substantial usability issues in one or both of their interactions (10 out of 35 participants). The replication showed similar patterns as the analysis of the entire data set. It was therefore concluded that any experienced interaction problems were unlikely to have substantially biased the findings.

Fig. 1.

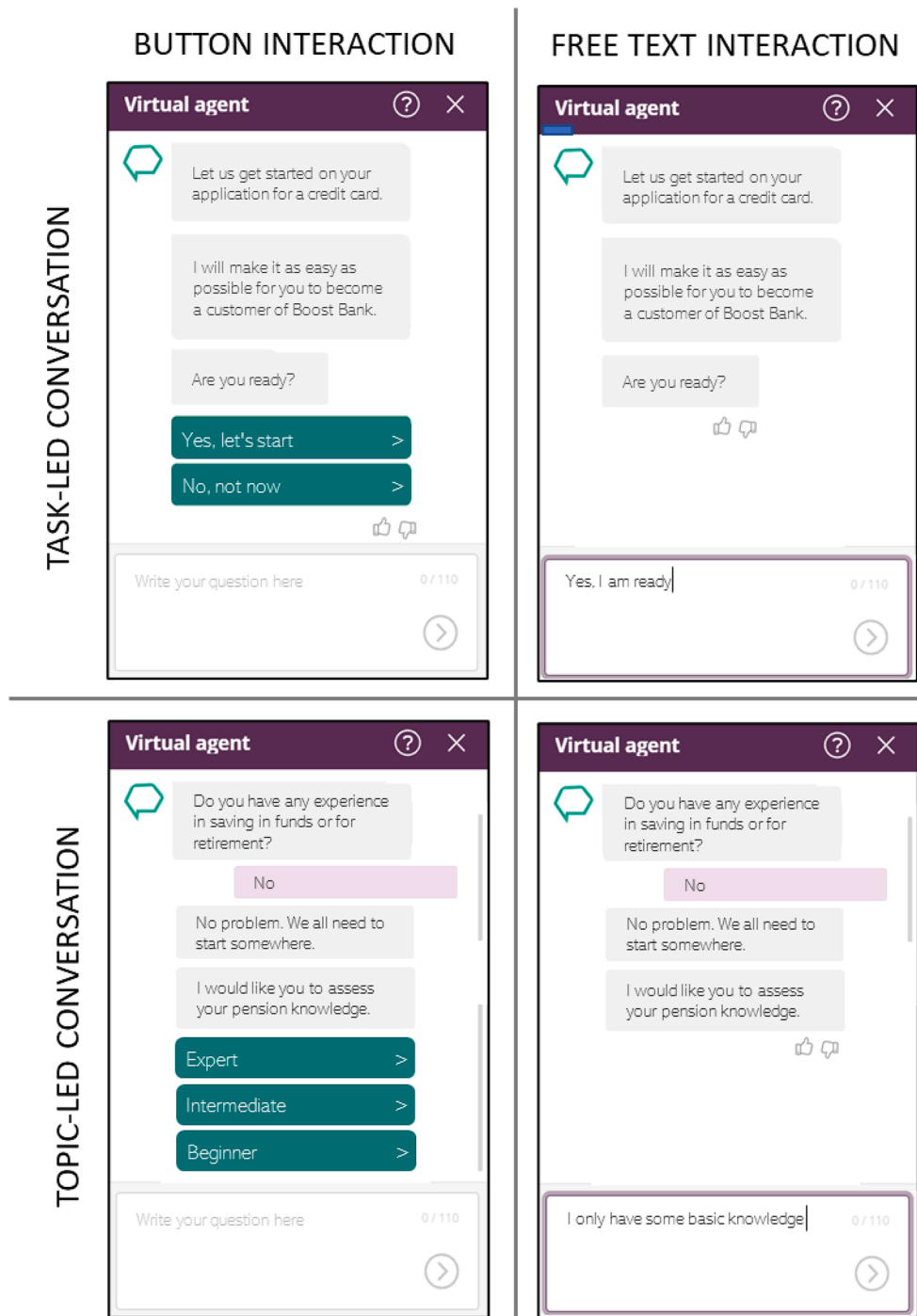[2] The participant recruitment agency was Norstat, https://norstat.no/

Fig. 1. Illustrations exemplifying the four chatbot conditions (texts in illustrations translated to English by the second author).

5. Results

In this section, we first detail findings concerning chatbot conversation types, then findings concerning interaction mechanisms. For the each of these, we first present findings from the measurement instruments and results of the statistical hypothesis testing before presenting the exploratory findings based on the interviews. The exploratory findings are initially presented with regard to perceptions of pragmatic and hedonic quality. Toward the end of the results section, we summarize exploratory findings concerning perceptions of human likeness and social presence.

5.1. The impact of conversation type

The impact of conversation type was measured for each dependent variable by way of the questionnaire scales and analysed through paired-samples t-tests. Conversation type was found to significantly impact the dependent variables in accordance with the stated hypotheses (H1-H3), with the exception of social presence (H1b) for which no significant effect was found.

Scores for perceived anthropomorphism (H1a) and hedonic quality (H2) were significantly higher for topic-led conversations than for task-led conversations. Furthermore, scores for pragmatic quality (H3) were significantly higher for task-led conversations than for topic-led conversations. Estimated effect sizes for these three dependent variables

were small to moderate following Cohen's rules of thumb (d = 0.36–.52).

The scores for social presence (H1b) were not significantly impacted by conversation type; the mean score was only slightly higher for topic-led conversations than for task-led conversations and a negligible effect size was estimated (d = 0.10).

The paired differences for the tested variables were not found to deviate from normality in a Shapiro-Wilks test. However, some outliers were found in their distributions. Because of this, the hypotheses tests were repeated by the non-parametric Wilcoxon signed ranks test. The analysis findings were largely replicated, though the difference between the conditions for anthropomorphism only bordered significance in the non-parametric test.

Details on the paired-samples t-tests are provided in Table 1.

Based on the interview data, our thematic analysis led to relevant insights on how conversation types may impact user experience. The user reports concerned aspects contributing positively to user experience as well as aspects contributing negatively. To provide a balanced and comprehensive overview of the analysis, we structured the themes in terms of aspects contributing positively and negatively to hedonic and pragmatic quality. We make particular note where the aspects concern user perceptions of anthropomorphism and social presence.

5.1.1. User perceptions of task-led conversations

In the interviews, task-led chatbot conversations were reported to entail aspects contributing substantially to pragmatic quality. In particular, the users reported to appreciate the conversation being *efficient and relevant* (28). The users noted the goal of the conversation to be clear and to experience the chatbot dialogue to rapidly take them through the needed steps.

> The chatbot did what I wanted it to, and it did so in a fast and efficient manner [P5]

The chatbot conversation style was reported to fit the task (20), that is, to keep with a professional and business-like manner well suited to a goal-oriented task. It was also interesting to note that most of the participants explicitly expressed preference for button interaction for the task-led conversation (19).

> It was very useful with alternatives provided, just because the objective was to submit a request […] this made the process easier [P17]

The participants also made note of hedonic aspects of the task-led conversations, though not as frequently as they did pragmatic aspects. These participants specifically noted the pleasure that may be felt by receiving efficient support from a chatbot (9). That is, they did not regard the smoothness of a goal-oriented process only with bland satisfaction but also associated it with a mild positive emotional experience.

> This was very smooth and seamless. It was efficient, so this gave a good feeling [P24]

Other participants reported on aspects of task-led conversations potentially reducing hedonic quality. Some (13) noted that they perceived the chatbot's task-led approach to be a bit too targeted and efficient, for example, as it immediately requested needed detail from the customer without any introductory politeness or pleasantries. While such a targeted approach is common in, for example, online forms, a chat interface may for some be expected to adhere more to conventions for humanlike conversational interaction.

Some (12) also noted the efficient character of the task-led conversation to make the chatbot seem more mechanical, predetermined or processual than what could be expected, for example, from a customer service employee. The efficiency which most participants appreciated may come at the cost of reducing the human likeness of the chatbot.

> Yes, it feels somewhat cold […] it was easy and fast, but not fun. You did not get any feeling of this being anything like a human [P26]

5.1.2. User perceptions of topic-led conversations

Participants reported on aspects concerning pragmatic quality also for the topic-led conversations, though positive reports on pragmatic quality were less frequent than for task-led conversations. This may not be surprising, as the objective of topic-led conversations is to trigger user reflection and engagement rather than achieving a specific goal. Some users reported that they appreciated a sense of progress or achievement in topic-led conversations (16), typically explaining that they did not see the point in conversing a chatbot unless they achieved something in the end.

> This was something resembling a dialogue, with sensible advancement or progress on a topic which I lack knowledge.[…] It was interesting to have a dialogue with. [P8]

Of relevance for our study of interaction mechanisms, it may be noted that some participants reported to find free text interaction more suitable for topic-led conversation (9). Specifically, these participants reported on the benefit to use their own words when the purpose of interaction was to explore a topic for which they lack knowledge.

> When you are to ask questions, it is clearly a benefit to be able to use your own words, as you then can ask the exact questions you want to [P17]

The majority of the participants reported on negative aspects of pragmatic quality when discussing the topic-led conversations. Specifically, the participants explained that they did not find the topic-led conversation sufficiently efficient or useful (23). That is, they did not immediately buy in to the notion of using the chatbot as a means of reflection and engagement.

> It was not really something wrong as such, but I feel that I did not get more from this than I would have gotten from reading up on the topic in the usual way [P6]

This sense of not immediately seeing the benefit of a chatbot conversation as a basis for reflection, also was expanded by some noting, in part, that the topic-led dialogue was not sufficiently interactive (7) and, in part, that the chatbot did not sufficiently adapt its content to the stated needs of the user (5). These issues may suggest that users possibly

Table 1
Results of paired-samples t-tests for the impact of conversation type on chatbot user experience (n = 35).

	Conversation type	*Mean*	*SD*	*t*	df	Sig. (2-tailed)	Effect size (*d*)
Anthropomorphism	Task-led	3.90	1.34	-2.15	34	< 0.05	0.37
	Topic-led	4.49	1.49				
Social presence	Task-led	4.46	1.61	-0.57	34	.58	0.10
	Topic-led	4.63	1.74				
Hedonic quality	Task-led	4.33	0.79	-3.12	34	< 0.01	0.52
	Topic-led	4.71	0.70				
Pragmatic quality	Task-led	5.53	1.27	2.10	34	< 0.05	0.36
	Topic-led	5.07	1.28				

expect more in terms of flexibility from a chatbot with topic-led conversations than from one with task-led conversations.

> The response to some of my input was not impressive. I guess I could have said a lot of different things and the chatbot had anyway proceeded with its story [P10]

The participants provided detailed insight into aspects which may cause topic-led conversations to be engaging or pleasurable. Most participants (20) noted the importance of the conversational content, explaining that well-crafted social content – such as pleasantries or humour – may contribute positively to the user experience. For example, an informal conversation style may make otherwise factual and dry content seem more engaging.

> It gives you more and is engaging [..] I get a sense of the chatbot understanding me when I say I have no experience, and I feel it is like a dialogue [P23]

About half the participants indicated topic-led conversations to strengthen perceptions of anthropomorphism (17). That is, the chatbot was seen as reflecting the character of a human conversational partner when inviting the user to informal reflections in a socially oriented language. Such human likeness was typically associated with engagement and a pleasurable user experience.

> Fascinating, so, yes, it is pleasant. A somewhat strange feeling of having a pleasant time with a machine. But fun. It was a positive experience. [P22]

However, many of the same participants also noted that the social content which typically is included in topic-led conversations may be challenging. Such content may require insight into how the communication is received by the conversational partner and how it fits with the context of the conversation – something that the chatbot may not have. Hence, while most participants made note of the possible pleasurable experiences of topic-led conversations, an equal proportion of the participants noted that the use of humour or pleasantries in a chatbot may

be problematic (23); it may be deemed unsuitable by some users or may not be sufficiently well timed.

> This may not be perceived the same way by all users. Some may like it, while others may see it as just additional text they have to read, which is not needed [P26]

Most users also noted that an informal conversation style may be seen as problematic by some users (21). While informal conversational content may possibly be seen as engaging, it may also introduce a risk of the chatbot being seen as less professional.

> It is more positive when it is not very personal. […] this is a bank, you know. I am not here to buy a hot dog [P3]

A summary overview of the themes concerning chatbot conversation types is provided in Table 2.

5.2. The impact of interaction mechanism

In the experimental setup, the type of interaction mechanism was found to significantly impact some of the dependent variables, but only partially in accordance with the stated hypotheses (H4-H6).

In line with our hypothesis (H6), button interaction was associated with significantly higher score for pragmatic quality than free text interaction. Estimated effect size for this difference was small to moderate following Cohen's rules of thumb ($d = 0.37$).

However, contrary to our hypothesis (H5), button interaction was also associated with significantly higher score for hedonic quality than free text interaction. Estimated effect size for this difference was also small to moderate following Cohen's rules of thumb ($d = 0.39$).

The scores for perceived anthropomorphism and social presence were, contrary to our hypotheses (4a and 4b), not significantly impacted by interaction mechanism.

The paired differences for the tested variables were not found to deviate from normality in a Shapiro-Wilks test. However, as some outliers were found in the distributions for hedonic quality and pragmatic

Table 2
Summary of themes from the analysis of participant reports on the studied conversation types.

Conversation type	Group of themes	Themes	Explanation
Task-led conversation	Pragmatic quality – positive perceptions	Efficient and relevant conversation (28)	Task-led chatbot reported to be perceived as efficient and relevant to the user goals and needs
		Tone of voice fit to task (20)	Conversational content in task-led chatbot reported to be a good fit for the task at hand
		Button interaction suitable (19)	Task-led chatbot reported to be particularly suited for button interaction
	Pragmatic quality – negative perceptions	(No relevant themes)	
	Hedonic quality – positive perceptions	Useful or efficient interaction is pleasurable (9)	Task-led chatbot reported to provide pleasant use experience through its efficient character
	Hedonic quality – negative perceptions	Overly efficient or targeted conversation (13)	Task-led chatbot reported to be overly efficient and goal-directed, which may reduce pleasure of interaction
		Mechanical interactions (12)	Task-led chatbot reported as overly schematic, which may reduce pleasure of interaction
Topic-led conversations	Pragmatic quality – positive perceptions	Sense of progress (16)	Topic-led chatbot reported to provide sense of productivity and progress during conversation
		Free text interaction suitable (9)	Topic-led chatbot reported to be particularly suited for free text interaction
	Pragmatic quality – negative perceptions	Not sufficiently useful or efficient (23)	Topic-led chatbot reported as not sufficiently useful or efficient, reducing perceived value
		Insufficient dialogue (7)	Topic-led chatbot not seen as dialogue-oriented, but rather a one-way communication
		Insufficient personalization of content (5)	Topic-led chatbot reported not to be sufficiently sensitive to context or personal preferences
	Hedonic quality – positive perceptions	Well-crafted social content (pleasantries and humour) (20)	Social content (e.g., pleasantries and humour) in topic-led chatbot reported to provide a pleasant experience.
		Topic-led chatbot humanlike (17)	Topic-led chatbot reported to have humanlike characteristics during the interaction
	Hedonic quality – negative perceptions	Ambiguous content, e.g., pleasantries or humour, (23)	Content intended as pleasantries and humour may lead to awkward communication experiences, e.g., if not well timed.
		Too informal conversational content (21)	Informal content or conversation style may be counter to expectations for professionalism in customer service.

Table 3
Results of statistical hypothesis tests for the impact of interaction mechanism on chatbot user experience (n = 35).

	Interaction mechanism	Mean	SD	t	df	Sig. (2-tailed)	Effect size (d)
Anthropomorphism	Buttons	4.39	1.33	1.39	34	.17	0.24
	Free text	4.00	1.53				
Social presence	Buttons	4.71	1.57	1.07	34	.29	0.18
	Free text	4.39	1.76				
Hedonic quality	Buttons	4.67	0.73	2.35	34	< 0.05	0.39
	Free text	4.37	0.78				
Pragmatic quality	Buttons	5.54	1.19	2.17	34	< 0.05	0.37
	Free text	5.06	1.35				

quality, the hypotheses tests were conducted also by the non-parametric Wilcoxon signed ranks test. The findings from the parametric tests were replicated in the non-parametric tests.

Details on the paired-samples t-tests are provided in Table 3.

The interview data provided additional insight into users' perceptions and experiences of the chatbot interaction mechanisms. In particular, this was valuable to explain findings in disagreement with our hypotheses. We first present the findings for button interaction, then for free text interaction. For each interaction mechanism, the identified themes are structured into aspects concerning pragmatic and hedonic quality respectively. We make note when these themes also concern reports indicating anthropomorphism and social presence.

5.2.1. User perceptions of button interaction

In the interviews, nearly all participants reported pragmatic aspects of button interaction to strengthen their user experience. Most reported button interaction to enable easy and efficient interaction (28) as users can choose from alternatives rather than type. Furthermore, about half (19) noted that button interaction may drive the conversation as the button alternatives help the user to stay on track towards their goal. This was in particular seen as beneficial for task-led conversations, but button interaction was also seen as beneficial for structure and progress in topic-led conversations.

> [With buttons] it is easier to decide, and also remember, what you want to ask. […] Hence, it will help you keep focussed, and it is helpful with this kind of information [P9]

Some pragmatic challenges were also reported with button interaction. Most participants noted that button interaction may be problematic if the alternative needed by the user is not available through the buttons (20). While some noted that in such cases one might just write the request in free text, rather than use the buttons, others found this potentially confusing or limiting.

> If someone thinks that none of these alternatives fit me it may be a bit … and if it is my grandfather and none of the button alternatives fit it may be difficult to understand that you can instead write something [P15]

A few participants also reported that button interaction may give users a sense of the chatbot being too much in charge of the conversation (6); that is, users' opportunities for affecting the process and outcome of the dialogue were seen as limited – something that may be counter to the notion of getting personal advice or to get help specifically tailored to users' contexts.

> When you use the button, you get a sense of somebody else have made all the choices for you [P33]

A few participants also made note of hedonic aspects of their button interaction. Some (7) pointed out different aspects of button interaction which may contribute positively to hedonic quality. For example, it was noted that the user may be more engaged when selecting from alternatives than when, for example, reading an informational text, that button interaction may increase trust as the user knows the conversation

will progress smoothly, or that selecting button alternatives may provide a comforting sense of privacy as the users do not have to specify their requests in their own words.

> It was very clear that 'here you can select' and here you can choose' […] I felt it comforting that it provided alternatives [P15]

A few participants also reported on aspects contributing negatively to hedonic quality. Some argued that button interaction may make the process seem less like an actual conversation (7) and more like a predefined dialogue flow. In consequence, button interaction may leave the user feeling as if selecting items from a predefined set of options. Some also argued that button interaction may make the conversation seem more mechanical or robotic (6) resulting in lower perceived anthropomorphism.

> The other felt more humanlike than this one […] You know, it feels less personal when there are answer alternatives [35]

5.2.2. User perceptions of free text interaction

When discussing free text interaction in the interviews, most participants reported that also this interaction mechanism may entail aspects contributing positively to pragmatic quality. Specifically, participants noted that free text interaction may strengthen flexibility (23), potentially paving the way for more personalized and relevant responses.

> It is positive that you can ask in a targeted way about exactly what you want and that you can control – at least so you believe – the dialogue in the direction you want, and get the answers you need [P8]

Furthermore, some participants reported that the free text interaction potentially provides the user with more direct control of the dialogue (12), which was seen as strengthening usefulness. That is, when the user is seen as more in command of the dialogue direction and outcome, it is seen as more likely that a relevant outcome will be achieved.

On the other hand, the participants also reported on a range of aspects contributing negatively to the pragmatic quality of free text interaction. Most participants argued that such interaction increases the chance of the chatbot misinterpreting users' intentions or requests (21), which was seen as potentially detrimental to usability. Hence, while free text may promise flexibility, it may also entail problematic issues unless natural language processing capabilities are sufficiently strong. Due to a fear of being misinterpreted, some participants also reported on a tendency to use simplified language in free text chatbot interaction (11), so as to increase chances of the chatbot understanding their request. Such simplification also holds pragmatic implications, as it may be more challenging to precisely describe a specific issue or area of interest.

> It may interpret what you have written in a way … if you have written it, sort of, with a typing error or an unclear formulation, it does not understand [P31]

A substantial number of participants (14) also reported free text

interaction pragmatically challenging as they found the typing to potentially be cumbersome, in particular when on smartphones, and that typing also could lead to misspellings which in turn could lead the chatbot to misinterpret their request. A few participants (8) also noted that free text interaction, as opposed to button interaction, was more cognitively demanding as they did not get any help from the answer alternatives to formulate their questions or requests.

> It is more cumbersome to write yourself. One has to put more thought into what to write compared to just selecting between different alternatives [P17]

Free text interaction was reported to hold important hedonic implications; some positive, others negative. Some participants reported free text interaction to contribute a pleasant experience as the interaction was seen as having a more personal feel (11). When being able to express themselves in their own words, these participants also found the chatbot to provide a more tailored or personalized experience.

> When you can ask your question in free text, you feel somewhat more seen or cared for [P3]

Some participants further reported the free text interaction to set the stage for a more humanlike or social interaction (9), which they found to strengthen the user experience.

> [With free text interaction] you are more part of the conversation. It is more humanlike and communicative [P26].

However, free text interaction may also entail aspects that reduce hedonic quality. Some participants expressed emotional distress when the free text interaction led to breakdowns in communication (11). That is, misinterpretations due to free text interaction was not seen as only holding pragmatic implications but also, for some, to be felt as deeply

frustrating.

Some also considered a chatbot's failure to interpret free text requests as a breach of expectations (6). That is, while free text interaction is seen as indicating an opportunity to ask just about any question within the chatbot's domain, limited interpretational capabilities in the chatbot may in practice cause users not to have more freedom than they would in a chatbot with button interaction. This tension between the promise of flexible free text interaction and the limited realities of interpretational capabilities may cause substantial negative impact on hedonic quality.

> In a way you are deceived to think that you have a kind of freedom that you do not really have after all. Because, if you write a lot, or something the chatbot does not know the answer to, you do not get an answer anyway. [...] I guess it is only when you chat with an actual person that it is better without answer alternatives [P15]

A summary overview of the themes concerning the studied interaction mechanisms is provided in Table 4.

5.3. General perceptions of anthropomorphism and social presence

In this final part of the results section, we present findings concerning the participants' general reflections concerning their perceptions of anthropomorphism and social presence in the chatbots. Table 5 provides an overview of the relevant themes. The themes are expanded below the table.

5.2.3. General perceptions of anthropomorphism

About half the participants (18) made specific note of their chatbot conversation feeling humanlike and personal, and most of this group considered this to contribute positively to user experience (15). When reflecting on their perceptions of anthropomorphism in the chatbots,

Table 4
Summary of themes from the analysis of participant reports on the studied interaction mechanisms.

Interaction mechanism	Group of themes	Themes	Explanation
Button interaction	Pragmatic quality – positive perceptions	Easy and efficient (28)	Button interaction reported to make conversation easy and efficient
		Drives the conversation (19)	Button interaction reported to help drive the conversation through clear dialogue paths
		Clarifies alternatives (16)	Available alternatives for answers and follow-ups are made clear through button interaction
	Pragmatic quality – negative perceptions	Alternatives may not be comprehensive (20)	Button interaction may be seen as counter productive if alternatives do not match user needs
		Leaves the chatbot in control (6)	Button interaction reported to leave the chatbot in charge of the dialogue, which may be disliked
	Hedonic quality – positive perceptions	Clear alternatives engaging or comforting (7)	Clear alternatives seen as strengthening hedonic aspects such as engagement, trust, or privacy
	Hedonic quality – negative perceptions	Less conversational experience (7)	Button interaction reported to entail reduced conversational experience
		Mechanical interaction (6)	Button interaction reported more schematic or mechanical, which may reduce engagement
Free text interaction	Positive pragmatic perceptions	Flexible interaction (23)	Free text interaction reported to allow for more flexible interactions with the chatbot
		Leaves user in control (12)	Free text interaction reported to leave the user more in control of the direction of the interaction
	Negative pragmatic perceptions	Interpretation issues (21)	Free text interaction reported to potentially entail interpretation issues in the chatbot
		Typing cumbersome and error-prone (14)	Typing free text questions may be strenuous or may entail the risk of typing errors
		Requires simplified language (11)	To avoid interpretation issues, free text interaction may require users to apply simplified language
		No help formulating requests (8)	In free text interaction, user provided no support in formulating alternative questions or refinements
	Positive hedonic perceptions	Personalized (11)	Free text interaction may give the user a sense of personalization or fit to own needs
		Anthropomorphic and social (9)	Free text interaction may give the user a sense of chatbot as more humanlike and socially oriented
	Negative hedonic perceptions	Incomprehension frustrating (11)	Free text interaction reported to lead to chatbot misinterpretation, considered frustrating
		Annoying if inflexible (6)	Inflexible free text interaction may be perceived annoying, may breach user expectations

Table 5
Overview of themes from the analysis of participant general reports on human likeness and social presence.

Group of themes	Themes	Explanation
Anthropomorphism – positive	Chatbot conversation personal or humanlike (18)	Participants reporting that conversation with a chatbot may feel like interacting with another human
	Human likeness beneficial (15)	Human likeness in chatbot interaction reported as beneficial for various aspects of the conversation, typically associated with hedonic quality.
Anthropomorphism – negative	Chatbots not seen as humanlike (15)	Chatbots reported not to be humanlike, typically associated with an argument that this is not possible for (current) machines
	Human likeness seen as artificial (11)	Attempts to make the chatbot overly humanlike, such as the use of humour, is reported to be artificial in an unpleasant way.
Social presence – positive	Politeness and pleasantries appreciated (14)	Socially oriented content in the chatbot, such as politeness and simple pleasantries, is reported to be appreciated even though it may not induce social presence.
Social presence – negative	Chatbots not seen as social or social aspects not important (22)	Chatbots reported not to be social, typically associated with an argument that this is not possible for (current) machines. Some also note that social aspects are not important for chatbots.

these participants noted that though they throughout the conversation were aware they interacted with a machine, they nevertheless found the chatbot human likeness to be pleasant, engaging, or amusing.

> The degree of human likeness is surprising … it is, I am all the time aware that this is a machine. That is, I do not feel a contact in that way, I just find it amusing. It does not replace a person, but it does the right things [P22]

The human likeness in the chatbot typically was reported to the be a consequence of the conversational content – such as the use of an informal conversation style, affective responses, humour, and emoticons – in addition to a less goal-oriented dialogue and choice of interaction mechanism, as treated above.

> It was really surprising. It was, all of a sudden, not so cold or automatic but, in a way, a sign of something human. Receiving positive feedback from a robot, it is sort of 'wow' [P20]

However, not all participants appreciated human likeness in chatbots. Nearly half (15) noted such human likeness to be unimportant. Furthermore, some (11) reported humanlike traits in chatbots to be artificial or unappealing. These participants argued this lack of appeal in human likeness to concern reduced efficiency in request handling. That is, conversational content intended for a humanlike experience could make interaction more extensive or cumbersome than needed.

> It was some humour, some anecdotes and some, which I find completely without value. I know that I am not talking to a human. Hence, this is completely unnecessary [P24]

Some of these participants also reported on a sense of unease or annoyance regarding humanlike chatbot characteristics. Specifically, these participants argued that they found the chatbot to appear as something it was not, which in the worst case could be seen as

distasteful.

> [One of the chatbots] tried to cross the border and become a bit too humanlike. […] It tries a bit too hard to be your friend, and it may feel repulsive [P27]

5.2.4. General perceptions of social presence

When asked about the chatbots' social presence, most participants reported that they did not see these as social, or that they did not perceive any social presence of the chatbots (22). Some of these participants reported to find the question of social presence curious; they did not see the chatbots as social conversational partners but as tools for acquiring information.

> Interviewer: How would you describe the chatbots as social conversational partners?
>
> Participant: That was a strange question. They are not social conversational partners. They are pre-programmed, responding to keywords […] They are not humans talking. [P1]

The participants also noted that the lack of social presence in the chatbots could be due to a lack of flexibility and adaptability in the chatbot. That is, the restricted domain of the chatbot, as well as limitations in terms of the chatbots ability to fully adapt to users' requests was seen as potentially barring perceptions of social presence.

While these participants did not perceive their chatbot interaction as social, some reported seeing the chatbot as having human likeness. In such cases, the participants typically reflected on the chatbot's human likeness as a feature which may provide a more pleasant or engaging interaction, but without the capability of generating social presence. That is, the human likeness in the chatbot was seen more as a characteristic of conversational content, or as a design element, rather than something genuinely social.

> This was not social in my view. Even though it was fun that the person making this chat or bot intended it to be humorous. To be social, you need you need to see responses that are directly related to what you have written [P25]

As a final point, it may be noted that when discussing social presence in the chatbots, or the lack thereof, some participants reported that they appreciated the chatbot as polite or pleasant to communicate with (14).

> There is pleasant, friendly feedback […] This is positive but means very little compared to user friendliness [P31]

Such characteristics suggest that well-crafted pleasantness in the chatbot mimicking aspects of social interaction may possibly contribute positively to user experience, though it may not induce a sense of social presence.

6. Discussion

In this section, we discuss our findings relative to the state of the art before summarizing implications for theory and practice, addressing study limitations, and suggesting avenues for future research.

6.1. How chatbot interaction design may impact user experience

Our findings shed new light on how chatbot interaction design may impact the user experience of chatbots for customer service. In particular, the findings suggest how the application of specific interaction design elements may impact hedonic quality and thereby user experience.

6.1.1. The impact of conversation types

While efficient problem resolution through task-led conversations is

key to any chatbot for customer service (Gartner, 2019; Kvale et al., 2020), topic-led conversations have been seen as promising for strengthening user engagement (Chung et al., 2020; a;Roller et al., 2020b; Shevat, 2017).

In line with our hypotheses, we found topic-led conversations to strengthen perceived anthropomorphism and hedonic quality in our chatbots. A similar finding was made in the participant interview reports. As such, the human likeness induced by a successful topic-led conversation may be seen as a providing something extra that is appreciated by the user and, hence, contribute a beneficial affective aspect to customer service (Berry et al., 2002).

However, while humanlike conversations for topical exploration may be seen as desirable by users, it is important to recognize that such use of humanlike design cues in chatbots may easily be overdone. Some of the participants noted that social content in customer service chatbots, such as humour and pleasantries may be seen as ambiguous or insensitive to the conversational context, potentially leaving users annoyed. The divergence in the participant interview reports on this issue suggests marked individual differences in how such social content in chatbots is received. In consequence, it will be important to make sure that the social content does not increase what Dixon et al. (2010) refers to as customer effort. This is also echoed in the participant interview reports, where it was noted that the pragmatic benefit of topic-oriented conversations depends on these being seen as progressing towards a desirable outcome. When interacting with chatbots for customer service, the user will hardly engage for the sake of conversation itself. A pragmatic outcome of the chatbot interaction is of paramount importance, also in topic-led conversations.

When this is said, also conversations aiming for highly pragmatic outcomes may benefit from introducing design elements increasing chatbot human likeness. From our findings, we particularly note that an overly targeted or efficient conversation – and the somewhat mechanical process of a highly goal-oriented chatbot – may work counter to hedonic quality. Hence, the same way skilled customer service personnel strengthen experience through responding to users in a courteous and pleasant manner (Hocutt et al., 2006) and adapt a fitting conversation style (Liebrecht and van Hooijdonk, 2019) a chatbot for customer service may strengthen user experience through careful application of pleasantries and politeness – even for task-led conversations.

Our findings clearly suggests that task-led conversations merit their prevalence in current customer service chatbots. At the same time, in line with the argument of Roller et al., 2020b on the benefit of in-depth conversations to convey knowledge and expertise, our findings indicate the potential of topic-led conversations in chatbots for customer service. However, to fully leverage topic-led conversations in chatbot design, future improvements in chatbot conversational capabilities are needed. As such conversational capabilities evolve, interaction designers will be able to flexibly apply both conversation types as part of customer service chatbots.

6.1.2. The impact of interaction mechanisms

In line with previous research (Jain et al., 2018; Valério et al., 2020), the findings from our hypotheses testing clearly demonstrate the pragmatic benefit of button interaction. Furthermore, the participant reports shed light on why button interaction has this beneficial effect on user experience; not only by providing a means of efficient interaction, but also by clarifying available alternatives to the user and driving the conversation along a desirable path for the user. As such, button interaction reduces cognitive load, minimizes chances of distractions, and helps users who lack the needed conceptual framework to ask the right questions and follow-ups. Button interaction was even by some reported to reflect hedonic quality, due to the engagement and comfort felt from being provided relevant alternatives.

Our findings did, however, not comply with our hypotheses for free text interaction. Chatbots with free text interaction scored significantly lower on both pragmatic and hedonic quality than did chatbots with

button interaction. Furthermore, no significant difference was found between chatbots with free text interaction and button interaction for perceived anthropomorphism and social presence. The participant interviews suggested several reasons for this. Free text interaction may entail interpretation issues, typing was seen as cumbersome and error-prone, and a lack of buttons was seen as receiving no support in formulating requests or follow-ups. Hence, even though free text interaction was seen as promising by some participants in terms of offering a more flexible interaction and leaving the user more in control, participants also questioned whether this flexibility was real as current chatbots for customer service may not have the conversational capabilities to back up promises of fully flexible interaction. Hence, our results expand on the findings of Jain et al. (2018), underscoring the importance of flexibility and conversational intelligence in the chatbot – which do require free text interaction – while acknowledging that free text interaction in itself may not be sufficient to strengthen user experience.

The availability of button interaction in chatbots is perceived to strengthen the pragmatic aspects of user experience and, hence, contributing to reduced customer effort (Dixon et al., 2010). However, button interaction may also be pragmatically limited. Users may want other response alternatives than what are offered on available buttons. For this, it will be valuable with an interaction design that clearly communicates the complementarity of free text interaction and button interaction.

6.1.3. Humanlike but not socially present?

Prior to our study, we assumed that strengthening the humanlike character of the chatbot – through choices in conversation types and interaction mechanisms – may strengthen perceptions of anthropomorphism and social presence. However, while topic-led conversations were found to strengthen perceived anthropomorphism, no such effect was found for social presence. This finding parallels Araujo (2018), who found humanlike design cues in the chatbot to positively impact anthropomorphism but not social presence. However, our finding is counter to the findings of Go and Sundar (2019), Schuetzler et al. (2018), and Diederich et al. (2019) who all found their manipulations of human likeness in the chatbot to impact both perceived anthropomorphism and social presence.

How may such differences in terms of perceived social presence between studies be explained? The chatbot interaction in our study, as in the study of Araujo (2018), was manipulated through global changes in conversation content. These changes affect the entire interaction, regardless of the users' input. Such global manipulations clearly may impact perceptions of anthropomorphism. For example, our participants report well-crafted social content in the conversations to strengthen the chatbot humanlike appearance. However, global manipulations of the chatbot dialogue may not be sufficient to impact perceptions of social presence. This is seen in our participants' reports of the chatbot interactions not being sufficiently flexible or adaptable to evoke a sense of immediacy and salience in the chatbot.

Previous studies that have identified effects for social presence all included manipulations more directly targeting the individual user than did the global manipulations in our study and in the study of Araujo (2018). Go and Sundar (2019), Schuetzler et al. (2018), and Diederich et al. (2019) all applied manipulations of chatbot interactivity which enhanced chatbot flexibility and adaptation to the individual user, rather than more global manipulations of chatbot content. Go and Sundar (2019) and Schuetzler et al. (2018) both applied techniques for backchannelling, that is, adapting the chatbot message content to reflect and repeat content from the users' messages. Diederich et al. (2019) adapted chatbot message content in response to analyses of sentiment in the users' messages.

Hence, while the perceptions of anthropomorphism may be strengthened by introducing global manipulations to the chatbot content, such as conversation type or introduction of humanlike design cues, inducing social presence may require stronger natural language

processing capabilities in the chatbot – such as capabilities for back-channelling or adaptations in response to user sentiment.

6.1.4. How conversation types and interaction mechanisms may be mutually beneficial

Our study was not designed to investigate interaction effects between variations in conversation types and interaction mechanisms. However, findings from the participant interview reports suggest that certain combinations of conversation types and interaction mechanisms may be particularly beneficial for enhancing pragmatic or hedonic quality. Specifically, some participants made note that the button interaction was particularly suitable for task-led interaction. Some participants also reported that free text interaction would be preferable for topic-led interaction. Such variation in fitness for purpose for the different interaction mechanisms is plausible also from indications in previous research. Jain et al. (2018) found users to appreciate the efficient task completion possible through button interaction, but also to be delighted by the conversational intelligence of chatbots with strong natural language capabilities. Similar assumptions are made in the interaction design recommendations of Shevat (2017), where task-led conversations are designed for efficiency while topic-led conversations are designed for engagement.

Clever use of button interaction to support free text requests and refinements likely is recommendable for task-led conversations. Furthermore, provided future chatbots for customer service achieve sufficiently strong natural language processing capabilities, free text interaction is likely required for sufficient flexibility in extensive topic-led interactions. Button interaction may be a valuable design element also for topic-led conversations in chatbots for customer service until sufficiently strong natural language processing capabilities are readily available.

6.2. Implications

6.2.1. Implications for theory

Topic-led conversations may strengthen perceived anthropomorphism and hedonic quality: Previous research has shown how different variations in conversational content may impact user experience, users' perceptions of anthropomorphism and, for some variations, also social presence (Araujo, 2018; Go and Sundar, 2019). Adding to this, our findings specifically demonstrate the importance of conversation type for user experience. Topic-led conversations may strengthen hedonic quality and perceived anthropomorphism in a chatbot, whereas task-led conversations may strengthen pragmatic quality, though at the possible cost of perceived anthropomorphism. Fully leveraging topic-led conversations in chatbot interaction design depends on strengthened conversation capabilities in chatbots. One path towards this end may be to provide extended support for identifying and addressing users main purpose of interaction, to complement current intent identification.

Button interaction may strengthen pragmatic and hedonic quality: Previous work (Jain et al., 2018; Valério et al., 2020) has shown the perceived benefits of button interaction for users, specifically for aspects of use associated with pragmatic user experience. Our findings verify and add to this by showing that button interaction may also, in some contexts, strengthen hedonic quality. For free text interaction, on the other hand, to strengthen user experience, our interview findings suggest that users will have to perceive the interaction as highly flexible and adapting to the personal needs of the user. That is, free text interaction in itself likely may not be a driver of strengthened hedonic quality. Such increase in hedonic quality may rather depend on underlying flexibility and adaptability which, in turn, requires free text interaction. Possibly, as future strengthening of conversational capabilities increasingly allows for more advanced topic-led conversations, free text interaction may more reliably indicate conversational capabilities and be used to strengthen chatbot human likeness.

Human likeness may not imply social presence: Previous

research has shown mixed evidence with regards to whether manipulations impacting chatbot human likeness also impacts social presence (Araujo, 2018; Diederich et al., 2019, Go and Sundar, 2019; Schuetzler et al., 2018). Our findings, like those of Araujo (2018), provide an example of a manipulation of conversational content which may impact perceptions of anthropomorphism without impacting social presence. It may be possible that such differences may be due to the manipulation having the character of a global changes to conversational content, such as changes in conversation type, and the character of local more flexible and personal adaptations within the conversation – for example through backchannelling (Go and Sundar, 2019; Schuetzler et al., 2018) or emotional awareness (Diederich et al., 2019).

6.2.2. Implications for practice

Strengthen hedonic user experience through topic-led conversations: Just like providing a pleasurable experience is important to customer service (Berry et al., 2002), strengthening hedonic quality in chatbots for customer service may improve user experience. Introducing topic-led conversations to the content of chatbots for customer service may have such an effect. Topic led conversations may strengthen user engagement as users are provided an interactional experience more reminiscent of interacting with another human. Users may in particular appreciate well-crafted social content as part of topic-led conversations, though it is important that such content fits the context of use and is robust to variation in user preferences. Furthermore, due to the productivity-focus in chatbots for customer service, it is important to craft topic-led conversations so as to have marked progress and lead to outcomes of direct value to the user.

Benefit from button interaction to strengthen user experience: Current chatbots for customer service typically allow users to write their inquiries in free text but do also provide button interaction for efficient dialogue management – for example, to help the user refine a request or to indicate relevant options (Shevat, 2017). Our findings verify the benefit of such button interaction to user experience. The findings also show that such interaction may also to some extent strengthen hedonic quality in addition to pragmatic quality. At the same time, service providers need to be aware of the limitations in button interaction – particularly the challenge experienced by users who do not find an adequate answer alternative amongst provided buttons. To mitigate this limitation, the design of the chatbots should make free text interaction a parallel and intuitive alternative for interaction whenever buttons are provided. Furthermore, it should be noted that future strengthening of conversational intelligence in chatbots for customer service may reduce the benefit in button interaction.

6.3. Limitations and future research

While the presented study has provided valuable insights into the interaction design of chatbots for customer service, it also has limitations. among these, we find three to be of particular importance.

First, the study followed a controlled laboratory research design where users interacted with chatbots as part of participating in the study. While this setup is a good choice to allow for investigating causality, it potentially threatens the ecological validity of the findings. Hence, to verify the generality of the findings also to real-world contexts we suggest future research to investigate the impact of conversation types and interaction mechanisms in the context of chatbots deployed as part of companies' service offerings. Such studies could, for example, be set up as A/B tests.

Second, the study sample was relatively small. Though adequate for the purpose of investigating the main effects of the studied factors, a larger sample size would be needed if the purpose was to study interaction effects. The findings based on the participant interviews suggested beneficial interactions between interaction mechanisms and conversation types. In consequence of this, we foresee future research with larger sample size to further investigate such interactions.

Third, the study context was that of customer service in the banking sector. Though we do not expect the chosen study context to be important for our findings it may nevertheless be valuable to replicate the findings also for chatbots for customer service within other sectors. Furthermore, to investigate the broader generality of the findings it could be interesting also to replicate the study outside the domain of customer service – for example in the context of chatbots for education or chatbots for health support.

Though limited, we see our study as providing a valuable contribution to the state of the art, while at the same time hopefully inspiring future research in response to the study limitations.

7. Conclusions

Our study findings provide new insights concerning the impact of conversation types and interaction mechanisms on chatbot user experience. Specifically, the findings provide insight into how introducing anthropomorphic elements into the interaction design of chatbots for customer service are perceived by users.

As expected, the use of a topic-led conversation type, designed to induce user reflection and engagement, may strengthen perceptions of human likeness and also of the hedonic quality in the chatbot. The participant interview reports suggest that human likeness in a chatbot indeed may impact hedonic quality as it facilitates more engaging and pleasant interactions. However, topic-led conversations may be seen as lower in pragmatic quality as these are less goal directed.

The choice of interaction mechanism was not found to impact user experience the way we expected. While button interaction indeed scored higher than free text interaction on pragmatic quality, free text interaction did not score higher than button interaction on perceived anthropomorphism or hedonic user experience. The participant reports suggest that this was due to the free text interaction not being associated with sufficiently flexible and adaptable conversational capabilities in the chatbot. Chatbot human likeness may not depend primarily on free text interaction, but rather on high levels of flexibility and adaptivity. Hence, free text interaction may then only be a necessary but insufficient prerequisite for this.

Finally, it may be noted that while human likeness may be feasible in current chatbots, even without advanced natural language processing capabilities such as backchannelling or adaptation to user sentiment, social presence may require such advanced capabilities. Users seem to acknowledge human likeness as a design element global to the entire interaction which may add positively to the overall user experience. However, they may not acknowledge the chatbot as a social conversational partner until the chatbot is able to adapt more flexibly to users' various individual needs and requirements. Social presence, hence, seem to require somewhat more advanced capabilities in chatbot interaction than what is available in regular intent-based chatbots for customer service. However, advances in chatbot research clearly may change this. Future advances in chatbot conversational intelligence, for example driven by the current progress in open domain chatbots (Adiwardana et al., 2020; Roller et al. 2020a), may entail exciting opportunities for strengthening user perceptions of anthropomorphism as well as social presence in chatbots for customer service, thereby strengthening the potential benefit of humanlike chatbot features such as topic-led conversations and free text interaction.

CRediT authorship contribution statement

Isabel Kathleen Fornell Haugeland: Conceptualization, Methodology, Data curation, Formal analysis, Resources, Writing – original draft. **Asbjørn Følstad:** Conceptualization, Methodology, Data curation, Formal analysis, Resources, Writing – original draft, Supervision, Funding acquisition. **Cameron Taylor:** Conceptualization, Resources, Writing – review & editing. **Cato Alexander Bjørkli:** Methodology, Supervision, Writing – review & editing.

Declaration of Competing Interest

We hereby confirm that the authors have no conflicts of interest of relevance for the study and conclusions presented in the manuscript.

Acknowledgement

Appendix A

The following questionnaire items and interview questions were applied in the presented study.

Questionnaire items

The questionnaire included measurement instruments for pragmatic and hedonic quality, human likeness, and social presence. All items of the measurement instruments are presented in the table below. Users responded to the items in Norwegian, translated by the authors.

Pragmatic Quality	Hedonic quality	Human likeness	Social presence
Semantic differentials, seven-point scale. Items: Complicated – Simple Impractical – Practical Cumbersome – Straightforward Unpredictable – Predictable Confusing – Clearly structured Unruly – Manageable	*Semantic differentials, seven-point scale* Items: Isolating – Connective Tacky – Stylish Cheap – Premium Alienating – Integrating Separates me – Brings me closer Unpresentable – Presentable Conventional – Inventive Unimaginative - Creative Cautious – Bold Conservative – Innovative Dull – Captivating Ordinary – Novel	*Semantic differentials, seven-point scale* Items: Machine-like – Human-like Unnatural – Natural Artificial – Lifelike	*Likert scales, seven-point scale, endpoints labelled disagree strongly and agree strongly* Items: I felt like I was engaged in an active dialogue with the chatbot. My interaction with the chatbot felt like a back-and-forth conversation. I felt as if the chatbot and I were involved in a mutual task. The chatbot was efficient in responding to my activities.
Adapted from Hassenzahl, Burmester and Koller (2003)	Adapted from Hassenzahl, Burmester and Koller (2003)	Adapted from Araujo (2018)	Adapted from Laban and Araujo, (2020)

Interview questions

Participant interviews were conducted (a) immediately after interaction with each chatbot condition, and (b) after both interactions were completed. The questions from the interview guide are presented in the table below.

Interview questions used immediately after interactions with Chatbot A and B respectively
1. How would you describe your experience of the chatbot you just used.
2. What are your thoughts on using the chatbot for this particular task?
Interview questions used after both chatbot interactions were completed
1. Please tell about your experience with Chatbot A
2. Please tell about your experience with Chatbot B
3. Please tell about similarities and differences between Chatbot A and Chatbot B.
4. How would you describe the usability of the two chatbots?
5. How would you describe your emotional experience of the two chatbots?
6. How would you describe the two chatbots as social communication partners?
7. How would you describe the two chatbots in terms of their human likeness?

References

Adam, M., Wessel, M., Benlian, A., 2020. AI-based chatbots in customer service and their effects on user compliance. Electron Mark. https://doi.org/10.1007/s12525-020-00414-7.

Adiwardana, D., Luong, M.T., So, D.R., Hall, J., Fiedel, N., Thoppilan, R., & Le, Q.V. (2020). Towards a human-like open-domain chatbot. arXiv preprint. arXiv: 2001.09977.

Araujo, T., 2018. Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. Comput. Hum. Behav. 85, 183–189.

Ashktorab, Z., Jain, M., Liao, Q.V., Weisz, J.D., 2019. Resilient chatbots: repair strategy preferences for conversational breakdowns. In: *Proceedings of CHI 2019* (paper no. 254). New York, NY. ACM.

Bernhaupt, R., Boldt, A., Mirlacher, T., Wilfinger, D., Tscheligi, M., 2007. Using emotion in games: emotional flowers. In: Proceedings of the International Conference on Advances in Computer Entertainment Technology. New York, NY. ACM, pp. 41–48.

Berry, L.L., Carbone, L.P., Haeckel, S.H., 2002. Managing the total customer experience. MIT Sloan Manag. Rev. 43 (3), 85–89.

Brandtzaeg, P.B., Følstad, A., 2017. Why people use chatbots. In: Proceedings of the International Conference on Internet Science – INSCI 2017. Cham, Switzerland. Springer, pp. 377–392.

Taylor, M.P., Jacobs, K., Subrahmanyam, K.V.J., Abirami, B., 2020. Smart talk. How organizations and consumers are embracing voice and chat assistants. Technical report. CapGemini.

Chen, H., Liu, X., Yin, D., Tang, J., 2017. A survey on dialogue systems: recent advances and new frontiers. ACM SigKDD Explor. Newslett. 19 (2), 25–35.

Chung, M., Ko, E., Joung, H., Kim, S.J., 2020. Chatbot e-service and customer satisfaction regarding luxury brands. J. Bus. Res. 117, 587–595.

Dale, R., 2016. The return of the chatbots. Nat. Lang. Eng. 22 (5), 811–817.

Diederich, S., Janssen-Müller, M., Brendel, A.B., Morana, S., 2019. Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: insights from an experiment with a conversational agent. In: *Proceedings of ICIS2019* (Paper no. 1560). AISEL.

Dixon, M., Freeman, K., Toman, N., 2010. Stop trying to delight your customers. Harv. Bus. Rev. 88 (7/8), 116–122.

Drift, 2018. The 2018 State of Chatbots Report. Technical report. Drift. Retrieved from. https://www.drift.com/blog/Chatbots-report/.

Følstad, A., Brandtzæg, P.B., 2017. Chatbots and the new world of HCI. Interactions 24 (4), 38–42.

Følstad, A., Brandtzaeg, P.B., 2020. Users' experiences with chatbots: findings from a questionnaire study. Qual. User Exp. 5 (3) https://doi.org/10.1007/s41233-020-00033-2.

Følstad, A., Skjuve, M., 2019. Chatbots for customer service: user experience and motivation. In: *Proceedings of Conversational User Interfaces – CUI 2019* (pp. 1-9). New York, NY. ACM.

Følstad, A., Kvale, K., Haugstveit, I.M., 2014. Customer support as a source of usability insight: why users call support after visiting self-service websites. In: Proceedings of NordiCHI 2014. New York, NY. ACM, pp. 167–170.

Forrester, 2016. The state of chatbots. Technical report. Forrester. Retrieved from. https://www.forrester.com/report/The+State+Of+Chatbots/-/E-RES136207.

Gartner, 2019. Market guide for virtual customer assistants. Technical report. Gartner. Retrieved from. https://www.gartner.com/en/documents/3947357/market-guide-for-virtual-customer-assistants.

Go, E., Sundar, S.S., 2019. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput. Hum. Behav. 97, 304–316.

Hassenzahl, M., 2018. The thing and I: understanding the relationship between user and product. In: Blythe, M., Monk, A. (Eds.), Funology 2: From Usability to Enjoyment. Springer International Publishing, Cham, Switzerland, pp. 301–313.

Hassenzahl, M., Burmester, M., Koller, F., 2003. AttrakDiff: ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Szwillus, G., Ziegler, J. (Eds.), *Mensch & Computer 2003: Interaktion in Bewegung*. Vieweg+ Teubner Verlag, Wiesbaden, Germany, pp. 187–196.

Hocutt, M.A., Bowers, M.R., Donavan, D.T., 2006. The art of service recovery: fact or fiction? J. Serv. Mark. 20 (3), 199–207.

ISO (2019). ISO 9214-210:2019 (E). Ergonomics of human system interaction-Part 210: human-centred design for interactive systems. International Organization for Standardization (ISO) Retrieved from: https://www.iso.org/obp/ui/#iso:std:iso :9241:-210:ed-2:v1:en.

Jain, M., Kumar, P., Kota, R., Patel, S.N., 2018. Evaluating and informing the design of chatbots. In: Proceedings of the Designing Interactive Systems Conference – DIS 2018. New York,NY. ACM, pp. 895–906.

Jenkins, M.-C., Churchill, R., Cox, S., Smith, D., 2007. Analysis of user interaction with service oriented chatbot systems. In: Proceedings of HCI 2007. Heidelberg, Germany. Springer, pp. 76–83.

Jurafsky, D., Martin, J.H., 2021. Chatbots & dialogue systems. In: Speech and Language Processing. Draft of December 29, 2021. Retrieved from https://web.stanford.edu/ ~jurafsky/slp3/24.pdf.

Kujala, S., Miron-Shatz, T., 2013. Emotions, experiences and usability in real-life mobile phone use. In: Proceedings of CHI 2013. New York, NY. ACM, pp. 1061–1070.

Kvale, K., Sell, O.A., Hodnebrog, S., Følstad, A., 2020. Improving conversations: Lessons learnt from manual analysis of chatbot dialogues. In: Proceedings of Conversations 2019–International Workshop on Chatbot Research and Design. Cham, Switzerland. Springer, Cham, Switzerland, pp. 187–200.

Laban, G., Araujo, T., 2020. Working together with conversational agents: The relationship of perceived cooperation with service performance evaluations. In: Proceedings of Conversations 2019–International Workshop on Chatbot Research and Design. Cham, Switzerland. Springer, Cham, Switzerland, pp. 215–228.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 159–174.

Laranjo, L., Dunn, A.G., Tong, H.L., Kocaballi, A.B., Chen, J., Bashir, R., Coiera, E., 2018. Conversational agents in healthcare: a systematic review. J. Am. Med. Inform. Assoc. 25 (9), 1248–1258.

Law, E.L.C., Van Schaik, P., 2010. Modelling user experience–an agenda for research and practice. Interact. Comput. 22 (5), 313–322.

Lee, S., Choi, J., 2017. Enhancing user experience with conversational agent for movie recommendation: effects of self-disclosure and reciprocity. Int. J. Hum. Comput. Stud. 103, 95–105.

Li, C.H., Yeh, S.F., Chang, T.J., Tsai, M.H., Chen, K., Chang, Y.J., 2020. A conversation analysis of non-progress and coping strategies with a banking task-oriented chatbot. In: Proceedings of CHI 2020 (paper no. 82). New York, NY. ACM.

Liao, Q.V., Mas-ud Hussain, M., Chandar, P., Davis, M., Khazaeni, Y., Crasso, M.P., Geyer, W., 2018. All Work and No Play?. In: Proceedings of CHI 2018 (paper no. 3). New York, NY. ACM.

Liebrecht, C., van Hooijdonk, C., 2019. Creating humanlike chatbots: what chatbot developers could learn from webcare employees in adopting a conversational human voice. In: Proceedings of CONVERSATIONS 2019–International Workshop on Chatbot Research and Design. Cham, Switzerland. Springer, pp. 51–64.

Meuter, M.L., Bitner, M.J., Ostrom, A.L., Brown, S.W., 2005. Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies. J. Mark. 69 (2), 61–83.

Meyer von Wolff, R., Hobert, S., Schumann, M., 2019. How may I help you? State of the art and open research questions for chatbots at the digital workplace. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. ScholarSpace, University of Hawaii, pp. 95–104.

Moore, R.J., Arar, R., 2018. Conversational UX design: an introduction. Studies in Conversational UX Design. Springer, Cham, Switzerland (pp. 1-16).

Nass, C., Moon, Y., 2000. Machines and mindlessness: Social responses to computers. J. Soc. Issues 56 (1), 81–103.

Nordheim, C.B., Følstad, A., Bjørkli, C.A., 2019. An initial model of trust in chatbots for customer service—findings from a questionnaire study. Interact. Comput. 31 (3), 317–335.

Palmer, A., 2010. Customer experience management: a critical review of an emerging idea. J. Serv. Mark. 24 (3), 196–208.

Papachristos, E., 2019. Assessing the performance of short multi-item questionnaires in aesthetic evaluation of websites. Behav. Inf. Technol. 38 (5), 469–485.

Parasuraman, A., Berry, L.L., Zeithaml, V.A., 1991. Perceived service quality as a customer-based performance measure: an empirical examination of organizational barriers using an extended service quality model. Hum. Resour. Manage. 30 (3), 335–364.

Pérez, J.Q., Daradoumis, T., Puig, J.M.M., 2020. Rediscovering the use of chatbots in education: a systematic literature review. Comput. Appl. Eng. Education 28 (6), 1549–1565.

Ram, A., Prasad, R., Khatri, C., Venkatesh, A., Gabriel, R., Liu, Q., & King, E. (2018). Conversational ai: The science behind the Alexa prize. arXiv preprint. arXiv: 1801.03604.

Roller, S., Boureau, Y.L., Weston, J., Bordes, A., Dinan, E., Fan, A., Williamson, M., 2020a. arXiv preprint.

Roller, S., Dinan, E., Goyal, N., Ju, D., Williamson, M., Liu, Y., & Boureau, Y.L. (2020). Recipes for building an open-domain chatbot. arXiv preprint. arXiv:2004.13637.

Schuetzler, R.M., Grimes, G.M., Giboney, J.S., 2018. An investigation of conversational agent relevance, presence, and engagement. In: Proceedings of ACIS 2018. AISEL.

Schuetzler, R.M., Grimes, M., Giboney, J.S., Buckman, J., 2014. Facilitating natural conversational agent interactions: Lessons from a deception experiment. In: Information Systems and Quantitative Analysis Faculty Proceedings & Presentations. Retrieved from. In: https://digitalcommons.unomaha.edu/isqafacproc/16.

Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-experimental Designs For Generalized Causal Inference. Houghton Mifflin, Boston, MA.

Shevat, A., 2017. Designing bots: Creating Conversational Experiences. O'Reilly Media, Boston, MA.

Smestad, T.L., Volden, F., 2018. Chatbot personalities matters. Internet Science. INSCI 2018. Springer, Cham, Switzerland, pp. 170–181.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., Loggarakis, A., 2020. User experiences of social support from companion chatbots in everyday contexts: thematic analysis. J. Med. Internet Res. 22 (3), e16235.

Valério, F.A., Guimarães, T.G., Prates, R.O., Candello, H., 2017. Here's what I can do: chatbots' strategies to convey their features to users. In: Proceedings of IHC'17 (article no. 28). New York, NY. ACM.

Valério, F.A., Guimarães, T.G., Prates, R.O., Candello, H., 2020. Comparing users' perception of different chatbot interaction paradigms: a case study. In: Proceedings of IHC'20 (article no. 11). New York, NY. ACM.

van der Goot, M., Hafkamp, L., Dankfort, Z., 2020. Customer service chatbots: a qualitative interview study into customers' communication journey. In: Proceedings of CONVERSATIONS 2020. Cham, Switzerland. Springer.

Weizenbaum, J., 1966. ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9 (1), 36–45.

Wilson, H.J., Daugherty, P.R., 2018. Collaborative intelligence: humans and AI are joining forces. Harv. Bus. Rev. 96 (7/8), 2–11.

Xu, A., Liu, Z., Guo, Y., Sinha, V., Akkiraju, R., 2017. A new chatbot for customer service on social media. In: Proceedings of CHI 2017. New York, NY. ACM, pp. 3506–3510.