Understanding Affective Experiences With Conversational Agents

Xi Yang
Imperial College London
London, UK
x.yang15@imperial.ac.uk

Marco Aurisicchio
Imperial College London
London, UK
m.aurisicchio@imperial.ac.uk

Weston Baxter
Imperial College London
London, UK
weston.baxter@imperial.ac.uk

ABSTRACT

While previous studies of Conversational Agents (e.g. Siri, Google Assistant, Alexa and Cortana) have focused on evaluating usability and exploring capabilities of these systems, little work has examined users' affective experiences. In this paper we present a survey study with 171 participants to examine CA users' affective experiences. Specifically, we present four major usage scenarios, users' affective responses in these scenarios, and the factors which influenced the affective responses. We found that users' overall experience was positive with interest being the most salient positive emotion. Affective responses differed depending on the scenarios. Both pragmatic and hedonic qualities influenced affect. The factors underlying pragmatic quality are: helpfulness, proactivity, fluidity, seamlessness and responsiveness. The factors underlying hedonic quality are: comfort in human-machine conversation, pride of using cutting-edge technology, fun during use, perception of having a human-like assistant, concern about privacy and fear of causing distraction.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Conversational Agents; User Experience; Positive Design; Affect; Emotional Design

ACM Reference Format:
Xi Yang, Marco Aurisicchio, and Weston Baxter. 2019. Understanding Affective Experiences With Conversational Agents. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300772

CHI 2019, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300772

1 INTRODUCTION

Conversational Agents (CAs) have been widely discussed since the beginning of this century. At that point, much of the research was around creating and evaluating human-like conversations with machines [5–7, 49]. This helped inform how CAs might be designed for wide scale use. In the past few years, CAs have become more ubiquitous. They have become widely available on smart phones (e.g. Apple Siri, Google Assistant), and are increasingly integrated into homes (e.g. Amazon Echo and Google Home) and cars (e.g. Google Assistant integration with Hyundai). Consequently, research has shifted to focus on how CAs can and will fundamentally contribute to many of our daily interactions. Examples of such research include explorations of how CAs can be used in healthcare and problem solving [18, 41, 43, 45]. With CA use now more common for the average person in everyday life, more research is needed to understand the experiences that users are having.

Recent research has investigated users' experiences with CAs focusing on system capabilities and gaps between users' expectations contrasted with their actual experiences [33, 37, 53], which lends itself to understanding user satisfaction. Other work has evaluated user satisfaction in the context of specific tasks, e.g. searching for information [1, 30, 50], food tracking [19], or design [48]. The experience explored in these studies is typically one of usability. In addition, these studies are often conducted as lab experiments, which can limit generalisability to real life scenarios [30, 50]. There remains relatively little work about understanding the affective responses towards conversational agents in real world situations. The most relevant work has been found in [9, 29, 42], where the authors propose the focus on affective capabilities of CAs such as how to take into consideration users' emotions. However, a deeper investigation of affective experiences in real world scenarios is still missing.

Emotions are at the heart of human experience [15]. In this paper, we argue that understanding users' affective experience is crucial to designing compelling CAs. This helps designers and engineers understand affective responses so they can reinforce the factors that people enjoy and mitigate the factors preventing positive experiences. As we will argue in the paper, the affective responses to a conversational agent are often as important as many of the functional evaluations that users make of CAs. The focus on affect is consistent with creating and enjoying a human-like interaction with digital entities, which is the ultimate goal of designing CAs.

The present study contributes new understanding of affective experiences with conversational agents and of the effects of perceived hedonic and pragmatic qualities of the agents on affective responses. The following research questions were asked to achieve this aim:

RQ1: What experiences do users have with CAs?
RQ2: What are the affective responses to these experiences?
RQ3: What product qualities influence affective experiences and how?

2 BACKGROUND

Conversational Agents

Within the literature, the term CA is often used interchangeably with Personal Assistant, Virtual Assistant, Intelligent Assistant, or Digital Assistant. Despite different names, CAs are consistently interpreted as applications, which provide assistance to the user by "*answering questions in natural language, making recommendations, and performing actions*" [2]. Among the many features distinguishing CAs from traditional applications, two major ones are the voice interface and the dialogue system, which is usually referred to as conversation [10]. Often such voice-enabled conversation requires an automatic speech recogniser to transcribe users' voice as input and convey information to the user either through text or voice output [17]. Existing CAs (e.g. Siri or Google Assistant) also integrate with the Graphical User Interface (GUI) in order to provide users with rich and interactive content. Derived from the notion of "virtual assistant" [39], the primary goals of CAs are often to assist users in real time and understand the knowledge of the users over time to offer better support [10].

Previous research has argued that the conversational interface has advantages over a traditional GUI by having a more human-like interaction [35], and by easing complex tasks with filtered information [4]. These advantages give CAs great potential to expand their applications in various life domains and contribute to many of our daily experiences.

Within the HCI community, there has been a growing interest in people's experiences with CAs [9, 10, 33]. One stream of research has focused on identifying important issues for CAs design and guiding directions for future research. For example, Cohen et al. in [9] have outlined four requirements for future generations of conversational agents, including CAs being able to recognise users' attentions, actively fulfil those attentions, collaborate with people or other assistants and engage in the conversations. Another stream of research has focused on understanding user needs and evaluating user satisfaction [31, 33, 35, 37, 53]. For example, Zamora in [53] has identified the need for "*high performing, smart, seamless and personal*" agents. The reality of current agents, however, does not live up to these expectations, as for example Luger and Sellen indicated that there is still a "*gulf between user expectation and experience*"[33]. A third stream of research has focused on studying CAs in specific domains such as complex searches [1, 30, 50], food tracking [19], and personas in design [48]. These studies have shown that it is promising to use CAs in novel scenarios. For example, Graf et al. showed that users rated the experience of food tracking by means of bots higher than that through common applications [19]. Similarly, Vtyurina et al. observed that users were glad to use conversational agents for complex searches as long as their expectations for accuracy were satisfied [50]. These studies have also highlighted the need to further investigate the use of CAs in different scenarios.

The experiences explored in previous studies are typically usability-oriented. There is relatively little work on understanding the affective responses towards conversational agents. The most relevant work is reported in [9, 42] where the affective capabilities of CAs are proposed to be as important as practical capabilities. For example, study [42] examines the relationship between the personification and sociability of Alexa and their effect on user satisfaction. They have found that personification predicts a higher level of sociability, and that sociability is positively related with user satisfaction. Cohen et al. in [9] have proposed important emotion-related issues to be addressed in the design of future CA such as how a CA should track user emotions and react to them. The above studies have contributed to a better understanding of the affective capabilities of CAs and the influence of such capabilities on user satisfaction. However, the literature still shows no large-scale examination of users' affective experiences with CAs in their daily lives.

Affective Experiences

It has been argued that user experience should go beyond the task-oriented approach of traditional HCI and focus on hedonic aspects such as fun and pleasure [25]. The importance of affective experience is also seen in various User Experience (UX) models and frameworks such as Don Norman's emotional design [38], Jordan's four levels of pleasures [28] and Hassenzahl's experience design framework [21]. Inevitably, emotions are at the heart of human experience and thus the hedonic aspects of UX are the keys to creating delightful and compelling experiences with CAs. Affective experience is often evaluated by capturing people's emotional responses, which can, for example, be assessed using the Positive Affect Negative Affect Schedule (PANAS) [51]. This widely used and validated scale provides rich affective information including the nuances of positive and negative affect.

A holistic understanding of UX requires the enrichment of traditional quality morals with non-utilitarian concepts, resulting in two types of product quality, i.e. pragmatic and hedonic [20]. As described in [20], pragmatic qualities are related to the functionality and usability of the system, which often fulfil users' behavioural goals, whereas hedonic qualities are related to non-instrumental qualities such as aesthetics, innovativeness or originality, which often emphasise users' psychological well-being. As shown in [23, 26, 36], pragmatic and hedonic qualities can be assessed using the AttrakDiff questionnaire [24].

Informed by this literature, we have adopted the above methods to study the affective experiences of CAs. Specifically we used the PANAS scale [51] and AttrakDiff questionnaire [24] to measure affective experiences and the perceived pragmatic and hedonic qualities.

3 METHODS

To answer the research questions, a survey study was designed to collect data on: 1) user experiences with CAs; 2) affective responses; and 3) perceived product qualities.

User experiences with CAs were collected using the critical incident method, which requires users to report an experience that they have had [14]. This method has been widely adopted in studies such as [13, 23, 26, 34, 36, 46, 47] to collect user experiences. In these studies, participants were asked to recall and report an experience. This is usually followed by a list of open-ended questions asking for more details of the reported experience. As this method is often effective in capturing rich information about an experience such as contextual information [13, 14], it is usually employed in studies which aim to collect and analyse real life user experiences.

Affective responses were measured using the PANAS scale [51]. In the literature, this scale is often used together with the critical incident method to assess affective responses of a recent experience. For example, in the study of human psychological needs in satisfying events, Sheldon et al. [46] asked participants to report on a recent, satisfying life event, and to rate the event using the PANAS scale indicating "*the extent to which they felt each of the different moods during the event*". Similarly, this approach has been shown effective in examining user experience with interactive technologies [23, 26, 34, 36]. For example in study [26], Hassenzahl et al. firstly asked participants to recall and report a recent experience with an interactive technology such as computer or smartphone, and then they asked participants to relive the experience and rate it on a series of scales including PANAS.

Perceived product qualities were measured using the AttrakDiff questionnaire [24]. This questionnaire measures three types of product qualities, namely pragmatic quality, hedonic quality and the overall appeal. Similar to the PANAS scale, this questionnaire is also often used together with critical incident method. For example in studies [23, 26, 36] examining technology-involved user experiences, the AttrakDiff scale was used together with the critical incident method to measure the pragmatic and hedonic qualities of products.

A survey method was adopted to collect the data mentioned above, as it allowed us to gain a broad overview of user experiences and quickly gain insights into the state of current practice. The survey was conducted across June and July 2017 using the Qualtrics platform and distributed using two participant recruitment services: Amazon Mechanical Turk and CrowdFlower.

In the study design we identified two main requirements. First, the main interest was in examining affective experiences with CAs and thus we did not want to introduce variations in application into the analysis. For example, it is known that CAs have different strengths across applications. Alexa for instance may support better shopping experience on Amazon. Siri for instance may provide better support on iOS. Because of this, we decided to focus on one CA only in this study. Second, we were interested in accounting for the ubiquitous experiences that users increasingly have with CAs and thus we wanted to have a CA which was accessible through multiple types of devices. The Google Assistant (GA) was considered suitable to address these requirements because it was accessible through multiple types of devices and in various contexts. For example, it was available on both smartphones and home devices such as Google Home. It could also be used on other devices such as smartwatch or through in-car systems. Other CAs, however, did not provide such various access devices. For example, Alexa was often used on home devices without stand-alone smartphone application. As a result, our recruitment criteria consisted of having had recent experience of using the GA through any of its access devices.

Survey Design

The first section of the survey introduced the study to the participants. The second section contained a series of open-ended questions asking participants about the most recent experience with the GA. We explicitly made participants aware that the experience could be positive, negative or neutral. Participants first described their experience and then answered questions regarding specific details, such as the device they used, the social and situational context, and the factors users enjoyed or were challenged by while using the GA. The third section assessed users' affective responses using the PANAS scale [51] with 10 positive affect descriptors and 10 negative descriptors. On a scale from 1 (not at all) to 5 (extremely), participants were asked to indicate to what extent they experienced a particular emotion. The fourth section evaluated the perceived qualities of the GA using the abridged version of the AttrakDiff2 questionnaire [24]. It consists of 10 items in the form of semantic differential scale (from -3 to 3), including *confusing-structured*, *impractical-practical*, *unpredictable-predictable*, and *complicated-simple* for the pragmatic quality, *dull-captivating*, *tacky-stylish*, *cheap-premium*, *unimaginative-creative* for the hedonic quality, and *good-bad* and *beautiful-ugly* for the overall attraction. The fifth section collected participants' demographic data, and the duration and frequency of using the GA.

4 DATA ANALYSIS

Cleaning and Preparation

A total of 210 individuals participated in the survey. Given that some reported experiences did not provide sufficient information we had to clean the data. Insufficient descriptions included those that were very short, abstract or general. For example, one participant described "*I am happy to try*". This response was excluded as it gave little information for analysis. After this process, 178 valid responses were left.

We further assessed the data according to which device was used in the interaction. In some cases, participants reported more than one device being used. This may have happened because some interactions relied on multiple devices. For instance, users sometimes needed to use a smartphone application to use Google Home. In these cases, we coded the device according to the primary device used to access the GA. Other times, participants chose more than one device as they possibly ticked all the devices that they used in the past. To tackle this issue, we examined the reported experience to identify the primary device used in that report. We dropped the data entries from which we could not identify the device.

After this process, we were left with 171 valid responses. Table 1 summarises participants' information. In this study, we collected both quantitative and qualitative data. The following analysis is based on the cleaned data.

Table 1: Participant Information

Age	Average 29.7 (SD=9.4)
Gender	73% male; 27% female
Nationality	70% the Americas; 16% Europe; 12% Asia; 2% Africa
Education	19% postgraduate; 71% undergraduate; 10% high school
Duration of use	50% > 12 months; 40% 1-12 months; 10% < 1 month
Frequency of use in past week	19% > 7 times; 21% 4-7 times; 53% 1-3 times; 7% none

Quantitative Analysis

Quantitative data were collected on users' affective responses and the perceived qualities of the GA.

We calculated the positive affect (PA) and negative affect (NA) value by averaging the responses to the 10 associated descriptors on the PANAS scale [51]. Internal consistency for PA and NA were at the excellent level (Cronbach's α of .90 for PA and .94 for NA). Besides the two factor distinction, PA can be further partitioned into three components: joy (*excited*, *proud*, *enthusiastic*), interest (*interested*, *strong*, *determined*) and activation (*active*, *inspired*, *alert*, *attentive*) [12]. Based on this partition, we calculated the values for joy, interest, and activation by averaging the respective descriptors. Internal consistency was satisfactory again (Cronbach's α = .83; .72; .76 respectively). To examine the nuances of positive affect, we decided to include these three components into our further analysis. To decide whether negative affect needed partitioning, we conducted a principal component analysis. Our results did not suggest the need for further partitioning, which is consistent with the work of Hassenzahl et al. [23].

Values of the pragmatic quality, hedonic quality and attraction were computed similarly, by averaging the scores of the associated descriptors. Internal consistency was satisfactory for all measures (Cronbach's α = .74; .78; .93 respectively). A Pearson Correlation Coefficient was further calculated to assess the relationship between the perceived qualities and affective responses.

Qualitative Analysis

Qualitative data were collected on users' self-reported experiences with the GA. We analysed the qualitative data following the thematic analysis process outlined by Braun and Clarke in [3]. The data were coded in terms of: 1) use scenario (what was the GA used for); 2) context of use including situational context and social context (where did the experience happen; who was with the participant); 3) device (which device was used to access the GA); 4) positive factors (what factors contributed to positive affect); and 5) negative factors (what factors contributed to negative affect).

The overall analysis was iterative. To start with, the first author derived initial coding categories. Next, all authors met to review the categorization, discuss specific data, and refine the categorisation. Following the final classification, two raters assessed a 10% random sample of the data to evaluate the consistency of the final coding scheme. Krippendorff's α [32] and Cohen's $\varkappa$ [8] were both calculated to validate the inter-rater consistency for the coding categories. Inter-rater agreement was satisfactory to all categories (Krippendorff's α ranging from 0.69 to 1; Cohen's $\varkappa$ ranging from 0.68 to 1). After the agreement was met, the remaining data were coded by the first author based on the developed categories.

5 RESULTS

In this section, we present results in response to the three research questions, i.e. users' experiences with the GA, their affective responses, and the product qualities that influence users' affective experiences.

User Experience

Scenarios of Use. The thematic analysis has shown four main scenarios of use: S1) requesting basic information; S2) searching for answers; S3) getting recommendations; and S4) accessing external services. While the first three scenarios are all related to information searches, they differ in terms of how users asked their questions, what they expected from the response, and whether there were follow-up interactions. They also differ in terms of the context of use such as the access device or location.

S1: Requesting Basic Information: In this scenario participants used the GA to request basic information. For example, participants used the GA to "*ask for weather*", "*find locations*", check "*news*" or "*sports updates*", and search for a "*player's name*", "*information about a band*", the "*population of a city*" and the "*definition of a slang term*".

The requested information was often a specific fact, and users were generally clear about the information they needed and how to ask for it. It was also easy for users to assess the relevance of answers. As a result, users usually received the expected answers without having to ask follow-up queries.

Our data showed that participants expected an accurate output. Moreover, participants enjoyed the experience when the GA proactively provided relevant information without being explicitly asked, for example, a participant stated "*I was getting ready for work … I pulled up the Google Assistant and wished it good morning. In response, it wished me a good morning too. Then it proceeded to tell me the weather, give me details about my commute and play the news program that I had set up. I enjoyed the fact that it gave me all the information that I wanted without me having to pull up a dozen different apps. I can just say 'good morning' and it tells me all that information while I'm getting dressed*" (P81).

In this scenario the response was typically presented as text extracts sometimes with images or web-links when necessary. We found that reading out the answers was generally preferred by participants, for example, "*I like making it read news to me when I was tired of reading texts*" (P59). However, it was notable that if the answers were not accurate enough, the reading-out feature could cause dissatisfaction. For example, a participant who was looking up information about a city's population reported: "*I used a voice command to access it … Of course, the first search result was a Wikipedia excerpt and it started to read it aloud. However, population was not in this excerpt. It was a bit annoying, so I had to open Wikipedia page and looked for it myself*" (P126).

S2: Searching for Answers: In this scenario participants used the GA to search for answers, which involved more complex information than the previous scenario and may require further processing of the search results. For example, a participant used the GA to learn about a plant: "*finding out the uses of a seed produced by a tree … and how can I turn it into a very good benefit for my home*" (P3), and another to learn about an astronomical event: "*looking up information about the recent eclipse of the sun … the interaction was very positive. It was suggesting what I may be interested in and this made my study fun*" (P109).

The requested information was more complex than facts and often required the user to assess its relevance by making judgments regarding inputs and search results. At times participants needed to alter the search query to receive better results, or to make follow-up queries.

Our data showed that to solve their problems participants expected an output that was useful. For example, one participant said, "*[the GA] gave me a useful response. I was satisfied*" (P102). Moreover, participants enjoyed it when the GA suggested relevant links for follow-up queries. For example, in the instance mentioned earlier where the participant was learning about the solar eclipse, he stated "*[the GA] was suggesting what I may be interested in*", and "*this made my study fun*" (P109).

In this scenario the response often involved the provision of links to webpages, images or videos. Hence, it was usually difficult for the GA to read out the answers directly.

S3: Getting Recommendations: In this scenario, participants used the GA to get recommendations on places to eat or visit. For example, participants asked for nearby locations such as the "*nearest restaurants*" (P11) or places with refined requirements "*a restaurant which serves Italian dishes near my location*" (P91). Other times, participants asked for travel advice: "*when travelling to a new place I ask the GA for place recommendations*" (P68).

The requested information was often a list of options satisfying users' requirements. Users then evaluated the options to make a decision. Therefore, there were follow-up actions, such as viewing restaurant details, calling for enquiries, making reservations, or getting directions to the place.

Our data showed that participants expected the output to include as much information as possible on the recommended places, for example, open hours, telephone or price. Such information was used to help them make decisions or with future actions. For example, a participant (P141) who "*asked for a Thai restaurant*" with the intention to "*order some* *takeout*" was expecting a contact number or order link from the response. Another participant shared an experience in which she was satisfied with the recommendation because she received the information she needed: "*I used it to find a coffee shop ... got all the information including price list very simply*" (P164).

In this scenario the response was often presented as a combination of text, images and sometimes locations.

S4: Accessing External Services: In this scenario, participants used the GA to access external services including home devices (e.g. interacting with lighting and TV), telephone services (e.g. making or receiving calls), media services (e.g. playing music), and personal task applications (e.g. checking calendar, reminders, and email).

Differently from the previous three scenarios, users' focus was more on interaction with the services rather than conversation with the assistant. For example, "*I was in the bathroom and heard my phone's notification sound. When I came out (with) my hands wet, I said 'OK, Google. Open WhatsApp' and boom! … It was very handy.*" (P8) In this example, the GA was used to launch WhatsApp, and there was no further conversation between the GA and the user.

Unlike the previous scenarios, the output was often external services responding to the user. In this scenario participants expected accurate and efficient execution of the command. For example, a participant stated "*this system seems wonderful to me because I can interact with my electronic devices almost immediately*" (P6).

Context of Use. Contextual information about the four scenarios including the device used, the location of use and the social context is summarised in Figure 1. Smartphone was the dominant device used across all four scenarios. Google Home also had a considerable role in the study. It was used most frequently when *Accessing External Services* (30%) and least frequently when *Getting Recommendations* (10%). Smartwatch was not commonly used by participants, and most of its use happened when *Accessing External Services*.

The two dominant locations of use were the home and in transit. When *Requesting Basic Information*, the home and in transit took the majority of the cases. Whereas, in the other three scenarios, the workplace and public places appeared commonly as well. In particular, when *Searching for Answers* and *Getting Recommendations*, a substantial number of experiences happened at public places (17.5%). Finally, a conspicuous number of experiences in the last three scenarios happened in the workplace (10%).

The social context was rather similar in the four scenarios. Participants used the GA predominantly alone, followed by uses in the presence of family members and lastly with friends or colleagues.

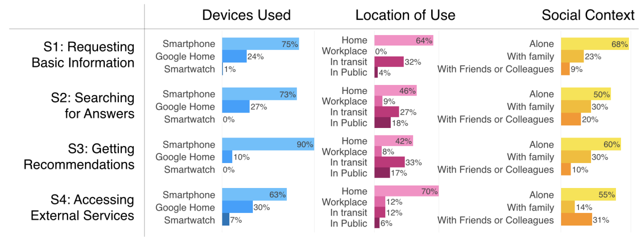

Figure 1: Context of Use Across Four Scenarios

Affective Responses

Participants' affective responses are shown in Table 2. The results showed that the overall positive affect was above 3 (moderately experienced), whereas the overall negative affect was lower than 2 (experienced a little). Interest was found to be the main contributor to positive affect, followed by joy and activation.

To test whether positive and negative affect were correlated with each other, e.g. if they were at the opposite poles of the same dimension, we performed the Pearson Correlation Coefficient calculation. The results showed no relationship between the two (r=0.07, p=.38). Therefore, positive and negative affect were found to be independent in our data, and were considered separately.

Table 2: Overall Evaluation of Affective Experience

Affect	Mean (SD)	Cronbach's α
Positive Affect	3.27 (0.93)	.90
Interest	3.38 (0.99)	.72
Joy	3.25 (1.14)	.83
Activation	3.12 (0.97)	.76
Negative Affect	1.80 (0.90)	.94

To investigate if the GA triggered different affective responses depending on the scenario, the five affect evaluators were calculated for each scenario. A normality test showed that the data were not all normally distributed. Therefore Kruskal-Wallis test was conducted. The results showed that there was a statistically significant difference for PA, interest and joy across the four scenarios ($\chi2(3,122)=8.18$, p=0.042; $\chi2(3,122)=8.32$, p=0.039; $\chi2(3,122)=8.40$, p=0.038 respectively), see Figure 2. The difference for activation and NA was, instead, found not to be statistically significant.

Following this, Mann-Whitney U tests were conducted to compare the differences of PA between pairs of scenarios. Significant difference was found between S1 and S2 (p=.02) and S1 and S4 (p=.03). This showed that *Requesting Basic Information* received significantly lower positive affect than

Searching for Answers and *Accessing External Services*. No significant difference was found between the other pairs.

The same tests were conducted for interest and joy. Significant difference was found between S1 and S2 (p=.02) and S2 and S3 (p=.03) for interest; and between S1 and S4 (p=.01) for joy. This showed that participants engaged in *Searching for Answers* experienced more interest than those engaged in *Requesting Basic Information* and *Getting Recommendations*, and that participants experienced more joy when engaged in *Accessing External Services* than in *Requesting Basic Information*. No significant difference was found between the other pairs.

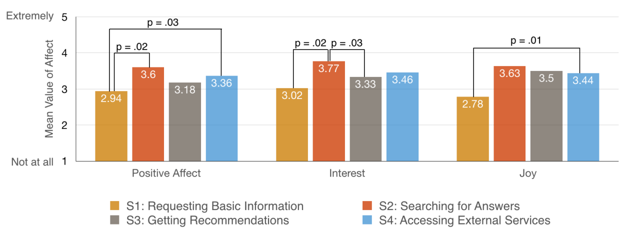

Figure 2: Affective Responses in the Four Scenarios

Product Qualities and Affect

Evaluation of product qualities showed that the overall attraction of the GA was rated the highest (M=1.66, SD=1.33) followed by the pragmatic quality (M=1.49, SD=0.98) and the hedonic quality (M=1.23, SD=1.22).

To test whether the evaluation of product qualities had effects on the affective responses, Pearson's r was calculated, see Table 3. According to the interpretation of correlation r [27], PA was positively correlated with all product qualities; NA was negatively correlated with all qualities except for the hedonic quality. This meant that on the one hand, increasing both the hedonic and pragmatic qualities could increase users' positive affect. Interestingly, the effect of the hedonic quality was higher than that of the pragmatic quality. On the other hand, among the two, only the pragmatic quality was found to significantly influence negative affect.

Table 3: Correlation Between Product Qualities and Affect

Product Quality	Positive Affect	Negative Affect
Hedonic	.47 *	-.07
Pragmatic	.32 *	-.42 *
Attraction	.28 *	-.35 *

* p < .001

In the next section, we elaborate on the product qualities by presenting the underlying factors identified from the thematic analysis.

Factors Underlying Pragmatic Qualities

Quality of the response content. The factors identified are *helpfulness* and *proactivity*.

Helpfulness, used here as an umbrella term, refers to the response being accurate, helpful or useful depending on the scenarios. For example, when *Requesting Basic Information*, participants used the term "*accurate*" to describe the response obtained. Whereas when *Searching for Answers*, participants used terms such as "*useful*" or "*helpful*" to describe how well the GA solved their problems. When *Getting Recommendations* participants used both "*accurate*" and "*helpful*" without a clear frequency difference between them. In addition, when *Accessing External Services*, participants appreciated how accurately the GA understood a command and how well the command was executed. For example, "*I enjoyed how accurately it understood me and executed the task*" (P150).

Inaccurate or irrelevant answers were found to cause negative experiences. For instance, a participant stated "*I asked it to help me find a location on a map. Results were not what I asked for so I was a little frustrated*" (P92).

Proactivity, that is predicting and responding to users' queries, was identified in the first three scenarios only. This may be because in these scenarios users regarded the GA as a smart search engine and thus expected more information. *Proactivity* was often realised by delivering additional information (i.e. information not directly requested) to the users. Such additional information was, however, expected to differ across the scenarios.

In *S1: Requesting Basic Information*, the additional information was typically offered based on contextual information and the results were not necessarily on the same topic. For example the participant getting ready for work said "*Good morning*" and was returned information about the "*weather*", the daily "*commute*", and the "*news*". These results were all relevant to the specific context of getting ready for work in the morning, but they covered diverse topics. In *S2: Searching for Answers*, additional information was typically provided on the same topic and was generated based on the search query. For example the participant who studied the solar eclipse was offered suggestions relevant to his subject of interest. In *S3: Getting Recommendations*, additional information was expected to be directly linked to the search results, for example, restaurant details such as price or reservation links. Generating such information may require the integration of results from different sources.

Quality of the interaction. The factors identified are: *fluidity*, i.e. maintaining a continuous voice interaction; *seamlessness*, i.e. integrating with external services; and *responsiveness*, i.e. shortening the time to obtain a response.

When fluid interaction was broken, a negative experience was likely to happen. For example, most of the frustration happened when the assistant could not understand users. This was often due to difficulty in recognising accent, lacking certain languages, or limited capabilities to understand complex sentences. Fluidity of interaction was often expected in multi-tasking contexts especially where hands-free was a must. It was also expected when the user was tired of "*typing or reading lengthy text*". In general, participants perceived the interaction with GA as "*easy to use*", "*convenient*", "*effortless*", "*intuitive*" and "*simple*". However, if the conversation was interrupted due to errors of voice interaction, negative feelings occurred. A robust and reliable technology behind voice interaction was shown to be the foundation for fluidity.

Seamless integration between the GA and external services was often appreciated by users. For example, the participants found the GA "*helpful for things like adding events to the calendar and even ask for an Uber car*" (P118). Another participant liked the GA because of the level of integration: "*I like it for its level of integration with all Google applications and the general search service.*" (P164)

Responsive interaction with the GA was valued as it was perceived saving time. Participants expressed their appreciation for this quality using terms such as "*so fast*", "*saves time and effort*", and "*immediately*". For example, a participant stated "*this system seems wonderful … because I can interact with my devices almost immediately*" (P6). A fast response was typically expected in all cases, but was especially appreciated when participants were in a rush or in trouble, for example, "*while driving in my regular route, one important road was blocked for some maintenance. [I] was not sure about the alternative route … I thought got stuck in some mystery place … the assistant said the way very perfectly along with the time to reach home. I had a very positive experience with this virtual assistant*" (P89). Our data showed that delays in the response were associated with negative affect. For example, a participant reported getting annoyed when the system was "*slow in answering*" (P78).

Factors Underlying Hedonic Qualities

Besides pragmatic qualities, our data also showed factors relating to hedonic qualities. These factors are: 1) comfort in human-machine conversation; 2) pride of using cutting-edge technology; 3) fun during interaction; 4) perception of having a human-like assistant; 5) concern about privacy; and 6) fear of causing distraction.

These factors usually do not directly affect the task at hand, which is the key difference between pragmatic and hedonic qualities. For example, errors during voice interaction may interrupt the task, whereas, uncanny feelings such as "*having to speak louder than normal*" (P100) were fine and did not stop the task. However, the hedonic factors were shown to influence how a user felt about the whole experience.

Lack of comfort was caused by the difference between a conversation with a machine and with a real person. It may be due to the way participants talked, for example, "*having to speak very clearly and loud*" (P157), or the way participants felt, for example, "*talking to no one feels like talking to self*" (P106) or "*cannot forget that I am talking to a robot*" (P29). Lack of comfort also happened if the user was not familiar with the GA, for example, "*I used it at my sister's house and it felt weird. I was used to talking to Siri so it was very different. I was uncomfortable because I wasn't used to it*" (P80).

For some participants, pride of using cutting-edge technology led to positive feelings. They expressed that using the GA was like "*experiencing the future*" (P29). The feeling of knowing, experiencing and learning new things made the participants proud, for example, "*I felt like I was modern. Showing it to someone who had never used it before made me feel like I know something that others don't*" (P158).

Besides saving time and being easy to interact with, the GA was, at times, perceived as "*fun to use*" (P63) and a device which "*breaks boredom*" (P21).

Furthermore, participants also expressed their connectedness to the GA, for example, "*I felt well taken care of*" (P71). In such cases, participants referred to the GA as a real person, for example, "*I feel someone is there to help me all the time. I never feel alone*" (P8), or that "*I felt very positive, like I have a personal assistant*" (P95).

Besides these factors, a few participants mentioned their concern for privacy, for example, "*the truth makes me afraid that Google knows every detail of our lives day by day*" (P162). Some other participants worried that interaction with the GA may be a distraction to the task at hand, especially in the multi-tasking contexts. For example, a participant mentioned "*I tried not to distract myself while driving*" (P58) and admitted not to like it.

6 DISCUSSION

This section discusses the main findings and the implications for existing and future CAs. The implications are not specific to a particular type of application, rather they are beneficial for positive CA design in general. For example, they could help designers better understand users' expectations across different scenarios and contexts, and therefore design for a positive user experience. More details on the findings and implications are discussed below.

User Experience

Scenario of use: more complex and diverse uses. Our study has identified four primary scenarios, which commonly appear among GA users. Consistently with previous findings [10, 33], the dominant scenario was shown to involve basic information requests such as weather updates or news. Our data also showed new and emerging scenarios which have become more common such as making more complex searches, getting place recommendations or accessing external services. The growing use of GA in complex and diverse scenarios may indicate that its reliability, familiarity and trust have increased. For example, study [33] stated that users were reluctant to use CAs for complex tasks or socially sensitive tasks such as making calls or sending long emails, and that users would not trust a CA in these circumstances. In our data, however, phone calls were one of the major uses, and users indicated enjoying the benefits brought by access to external services.

Context of use: more ubiquitous uses. As Fischer et al. state in [40], the use of CAs is expanding towards more social and collaborative contexts. Because of this, in HCI and CSCW it has been proposed to explore the use of CAs in everyday real-life settings. In agreement with Fischer et al., our study has shown growing ubiquity. Besides home and car being the dominant locations of use, workplaces and public settings such as street, park, or café have also been identified as common use settings. Our findings showed that public settings appear more frequently in scenarios where users ask for recommendations, for example, asking for places to eat or visit. Public settings also appeared frequently when people used the GA to solve their problems through more complex searches. The social context is expanding too, from private settings towards social settings with family, friends or colleagues being present. Our data have shown that about half of the uses happened when users were alone; the other half were with people around. In line with [9, 40], the evidence that the social context is evolving sheds new light on the diversity and ubiquity of GA use, and invites future research to study conversations within multi-user settings.

Implications. In his experience design theory, Hassenzahl has proposed that design should start with *why* (i.e. clarifying the needs and emotions involved in an experience), and then proceed with *what* (i.e. functionality) and *how* (i.e. ways to realise the functionality) [21]. The scenarios and contexts identified in this research together with the associated emotions can help designers understand the *why* of CA experience. Specifically, designers could use the types of scenarios and contexts emerged in this study to understand the needs of users. For example, our findings have shown that when *Requesting Basic Information*, users expect the response to include additional information which could be helpful for the activity that they are engaged in.

Affective Responses

Our findings showed that GA users experienced more positive than negative affect. This finding, however, disagrees with results from previous studies [1, 33, 50], which stated that a big gap existed between user expectations and the practical realities of use, and that satisfaction was generally low. This contradiction may be explained in two ways. First, CAs have significantly developed and improved. By the time previous studies were conducted, the technology investigated in this research was not as developed as it is now. The present study (conducted in 2017) may have benefited from a more reliable technology support, thus obtaining more positive experiences. Second, users' familiarity with the GA has increased. Therefore, users are more aware of what the agent can offer and what to expect from it.

Further study of the sub-components of positive affect has shown that interest is the most salient emotion experienced by participants, followed by joy and activation. This may be explained by the emotional design theory [52] which states that interest is stimulated by the appraisal of novelty-complexity ("*a product must be appraised as novel and/or complex to be interesting*") and coping potential ("*the degree to which one appraises oneself to have sufficient skills, knowledge, and resources to deal with an event*"). In the context of the present study, it is likely that the GA was considered as a novel product, due to the young age of CAs in general. This may also be because the natural language interaction has made it so intuitive to use that the required coping skills are generally low, and thus the coping potential is high. Novelty as a motivation to use CAs has been found in recent research [33]. For example, people were found to use CAs because they are curious, or they simply want to play with them. However, it is worth to note that such novelty or curiosity may decrease when people become more familiar with applications, thus causing a loss of interest [52]. Such effect is already evidenced in our data, as participants engaged in requests for basic information received significantly lower interest than those engaged in more complex searches.

Implications. Two approaches may be useful to maintain a positive experience level. The first approach involves increasing the complexity level of tasks, while keeping the coping potential high, for example, by introducing challenging tasks. This approach is grounded in the belief that increasing complexity could stimulate interest [52]. This is also evident in our finding that when *Searching for Answers* users experienced significantly higher level of interest than in the other scenarios. The second approach consists of increasing the level of joy and activation. This could for example be achieved by introducing more playful interactions between the product and the user [11, 16]. In our findings it was observed that users experienced significantly higher level of joy when they used the GA to access external services compared to requesting information. This may indicate that joy was elicited through the interactive experience with external devices. Further, activation theory has shown that a certain level of activation is necessary for effective functioning and that such activation can be stimulated by novelty, complexity, variation and uncertainty [44].

Factors Influencing Affective Responses

The quantitative results have shown that increasing both pragmatic and hedonic qualities could significantly improve positive affect. This result is consistent with previous studies [23], where a correlation between product qualities and positive affect is identified for interactive technologies. Previous work has suggested that the design of interactive technologies could benefit from considering both pragmatic (task-related aspects) and hedonic (task-unrelated aspects) qualities [20]. This is because both aspects substantially contribute to the overall appeal of a system [22]. Our findings show that this suggestion is valid for the design of CAs as well.

Pragmatic and hedonic factors. With respect to the pragmatic factors, Zamora found that users expect a high-performing, smart and easy-to-use agent [53]. Our study showed that the GA has met some of these expectations. For example, high-performing was evidenced by the GA being helpful; and easy-to-use evidenced by its interaction being fluid, seamless and responsive. Proactivity of CAs has also been proposed in the literature [9]. In particular, future CAs have been proposed to be capable of recognising users attentions or goals and actively responding to such attentions. Again, our study has found such traits in the GA. However, more importantly, our study contributes to the literature by validating the link between these factors and affective responses. As shown in our findings, these pragmatic-related factors would significantly contribute to positive affect if satisfied, but would significantly cause negative affect if not. This suggests that reinforcing such factors in design would improve users' positive experiences and reduce negative ones.

Hedonic factors are less discussed in the CA literature compared to pragmatic ones. In [9], it has been asked whether and how a CA could recognise users' emotions; and if users can be in multiple emotional states simultaneously with varying degree of intensity. Our study contributes to this literature highlighting that users' emotions may differ across scenarios depending on the tasks that they perform and the contexts that they are in. Our findings also showed that multiple emotions could coexist with different intensity levels. Not only different types of positive emotions may coexist, but also positive and negative emotions could be present simultaneously. More importantly, our study contributes to the literature by validating the link between these factors and affective responses. As shown in our findings, hedonic-related factors such as *comfort*, *pride* and *fun* during use, or the *perception of having a real assistant* can significantly contribute to positive affect. Positive affect can also be improved if concerns for *privacy* or *distraction* are reduced.

Implications. The identified pragmatic and hedonic factors have the potential to be used as guidelines and benchmarks in the design and evaluation of experiences with CAs. Specifically, to provide helpful and proactive answers designers could offer additional information or links to help users proceed with their activities. To deliver fluid interactions designers could focus on enabling robust voice interaction as this can help reduce frustration due to weak comprehension. To seamlessly integrate CAs with external services, designers could anticipate users' preferences for external services and launch them accordingly. To increase the responsiveness of CAs designers could attempt to address the waiting time, for example, by explaining to users why an answer cannot be delivered immediately. With respect to hedonic factors, to reduce discomfort when talking to an agent designers could introduce onboarding processes or tutorials, for example, educating users that they do not need to speak loud or slow with agents. To introduce fun interactions designers could create opportunities to break boredom or make tasks more playful. To develop CAs with the traits of a human assistant designers could design CAs with caring personalities. Finally, to reduce concerns for privacy and distraction designers could attempt to increase user trust.

Limitation

This study adopted the critical incident method to collect users' self-reported experiences. One limitation of this approach is that recalled memories may be inconsistent with interactions observed in process. For the purpose of understanding the experiences of users this method is adequate since we are primarily concerned with subjective experiences as reported by users. Additionally, the method is typically criticised as participants may have not invested adequate time in reporting rich information. This limitation was mitigated by asking more in-depth questions in the survey requiring specific information about the reported experiences.

7 CONCLUSION

This study contributes to the literature on conversational agents by extending our understanding of user's affective experiences with the GA. Specifically, we performed a quantitative evaluation of the affective responses and product qualities. We also performed a qualitative analysis of users' experiences with the GA describing four primary use scenarios, and multiple factors behind the affective responses. We found that the overall experience was positive with interest being the dominant contributor followed by joy and activation. The GA has become more ubiquitous and compared to previous findings is used in increasingly diverse scenarios and contexts leading to different affective responses.

REFERENCES

[1] Sandeep Avula. 2017. Searchbots: Using Chatbots in Collaborative Information-seeking Tasks *(SIGIR '17)*. ACM, New York, NY, USA, 1375–1375. https://doi.org/10.1145/3077136.3084155

[2] C. Baber. 1993. *Developing interactive speech technology, Interactive speech technology: human factors issues in the application of speech input/output to computers.* Taylor & Francis, Inc., Bristol, PA.

[3] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa

[4] Susan E. Brennan. 1990. Conversation as direct manipulation: An iconoclastic view. (1990). http://www.citeulike.org/group/890/article/633531

[5] Justine Cassell. 2000. *Embodied conversational agents.* MIT press.

[6] Justine Cassell, Catherine Pelachaud, Norman Badler, Mark Steedman, Brett Achorn, Tripp Becket, Brett Douville, Scott Prevost, and Matthew Stone. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. ACM, 413–420.

[7] Justine Cassell and Kristinn R. Thorisson. 1999. The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence* 13, 4-5 (1999), 519–538.

[8] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 1 (1960), 37–46. http://journals.sagepub.com/doi/pdf/10.1177/001316446002000104

[9] Phil Cohen, Adam Cheyer, Eric Horvitz, Rana El Kaliouby, and Steve Whittaker. 2016. On the future of personal assistants. ACM, 1032–1037. http://dl.acm.org/citation.cfm?id=2886425

[10] Benjamin R. Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. What can i help you with?: infrequent users' experiences of intelligent personal assistants. ACM, 43. http://dl.acm.org/citation.cfm?id=3098539

[11] Pieter M.A. Desmet. 2012. Faces of Product Pleasure: 25 Positive Emotions in Human-Product Interactions. *International Journal of Dsign* 6, 2 (2012), 1–29. http://www.ijdesign.org/ojs/index.php/IJDesign/article/view/1190/459

[12] Boris Egloff, Stefan C. Schmukle, Lawrence R. Burns, Carl-Walter Kohlmann, and Michael Hock. 2003. Facets of dynamic positive affect: Differentiating joy, interest, and activation in the positive and negative affect schedule (PANAS). *Journal of personality and social psychology* 85, 3 (2003), 528–540. https://www.researchgate.net/profile/Boris_Egloff/publication/5995270_Facets_of_Dynamic_Positive_Affect_Differentiating_Joy_Interest_and_Activation_in_the_Positive_and_Negative_Affect_Schedule_PANAS/links/540982830cf2822fb738e692.pdf

[13] Malin Eiband, Mohamed Khamis, Emanuel von Zezschwitz, Heinrich Hussmann, and Florian Alt. 2017. Understanding shoulder surfing in the wild: Stories from users and observers. ACM, 4254–4265. http://dl.acm.org/citation.cfm?id=3025636

[14] John C. Flanagan. 1954. The critical incident technique. *Psychological bulletin* 51, 4 (1954), 327. http://psycnet.apa.org/journals/bul/51/4/327/

[15] Jodi Forlizzi and Katja Battarbee. 2004. Understanding experience in interactive systems. ACM, 261–268. http://dl.acm.org/citation.cfm?id=1013152

[16] Barbara L. Fredrickson, Michael A. Cohn, Kimberly A. Coffey, Jolynn Pek, and Sandra M. Finkel. 2008. Open hearts build lives: positive emotions, induced through loving-kindness meditation, build consequential personal resources. *Journal of personality and social psychology* 95, 5 (2008), 1045.

[17] James Glass. 1999. Challenges for spoken dialogue systems. http://www.cs.cmu.edu/afs/cs.cmu.edu/Web/People/dod/papers/glass99.pdf

[18] Arthur C. Graesser, Nia Dowell, and Danielle Clewley. 2017. Assessing Collaborative Problem Solving Through Conversational Agents. Springer, 65–80.

[19] Bettina Graf, Maike Krüger, Felix Müller, Alexander Ruhland, and Andrea Zech. 2015. Nombot: Simplify Food Tracking *(MUM '15)*. ACM, New York, NY, USA, 360–363. https://doi.org/10.1145/2836041.2841208

[20] Marc Hassenzahl. 2003. The thing and I: understanding the relationship between user and product. Springer, 31–42. https://link.springer.com/content/pdf/10.1007/1-4020-2967-5_4.pdf

[21] Marc Hassenzahl. 2010. *Experience Design: Technology for All the Right Reasons.* Morgan & Claypool Publishers.

[22] Marc Hassenzahl, Andreas Beu, and Michael Burmester. 2001. Engineering joy. *Ieee Software* 18, 1 (2001), 70. http://search.proquest.com/openview/5d220f6a29326527ad73de88fe581782/1?pq-origsite=gscholar

[23] Marc Hassenzahl, Sarah Diefenbach, and Anja Göritz. 2010. Needs, affect, and interactive products–Facets of user experience. *Interacting with computers* 22, 5 (2010), 353–362. http://www.sciencedirect.com/science/article/pii/S0953543810000366

[24] Marc Hassenzahl and Andrew Monk. 2010. The inference of perceived usability from beauty. *Human–Computer Interaction* 25, 3 (2010), 235–260. http://www.tandfonline.com/doi/abs/10.1080/07370024.2010.500139

[25] Marc Hassenzahl and Noam Tractinsky. 2006. User experience - a research agenda. *Behaviour & Information Technology* 25, 2 (March 2006), 91–97. https://doi.org/10.1080/01449290500330331

[26] Marc Hassenzahl, Annika Wiklund-Engblom, Anette Bengs, Susanne Hägglund, and Sarah Diefenbach. 2015. Experience-Oriented and Product-Oriented Evaluation: Psychological Need Fulfillment, Positive Affect, and Product Perception. *International Journal of Human-Computer Interaction* 31, 8 (Aug. 2015), 530–544. https://doi.org/10.1080/10447318.2015.1064664

[27] Ronald J. Hunt. 1986. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. *Journal of Dental Research* 65, 2 (1986), 128–130.

[28] Patrick W. Jordan. 2002. *Designing Pleasurable Products: An Introduction to the New Human Factors.* CRC Press.

[29] Malte F. Jung. 2017. Affective Grounding in Human-Robot Interaction. ACM, 263–273.

[30] Julia Kiseleva, Kyle Williams, Jiepu Jiang, Ahmed Hassan Awadallah, Aidan C. Crook, Imed Zitouni, and Tasos Anastasakos. 2016. Understanding user satisfaction with intelligent assistants. ACM, 121–130. http://dl.acm.org/citation.cfm?id=2854961

[31] Hannu Korhonen, Juha Arrasvuori, and Kaisa Väänänen-Vainio-Mattila. 2010. Analysing user experience of personal mobile products through contextual factors. ACM, 11. http://dl.acm.org/citation.cfm?id=1899486

[32] Klaus Krippendorff. 2007. Computing Krippendorff's alpha reliability. *Departmental papers (ASC)* (2007), 43. http://repository.upenn.edu/cgi/viewcontent.cgi?article=1043&context=asc_papers

[33] Ewa Luger and Abigail Sellen. 2016. Like having a really bad PA: the gulf between user expectation and experience of conversational agents. ACM, 5286–5297. http://dl.acm.org/citation.cfm?id=2858288

[34] Elisa D. Mekler and Kasper Hornbæk. 2016. Momentary Pleasure or Lasting Meaning?: Distinguishing Eudaimonic and Hedonic User Experiences *(CHI '16)*. ACM, New York, NY, USA, 4509–4520. https://doi.org/10.1145/2858036.2858225

[35] Robert J. Moore, Raphael Arar, Guang-Jie Ren, and Margaret H. Szymanski. 2017. Conversational UX Design *(CHI EA '17)*. ACM, New York, NY, USA, 492–497. https://doi.org/10.1145/3027063.3027077

[36] Livia J. Müller, Elisa D. Mekler, and Klaus Opwis. 2015. Facets In HCI: Towards Understanding Eudaimonic UX–Preliminary Findings. ACM, 2283–2288. http://dl.acm.org/citation.cfm?id=2732836

[37] Quynh N. Nguyen and Anna Sidorova. 2017. AI capabilities and user experiences: a comparative study of user reviews for assistant and non-assistant mobile apps. (2017). http://aisel.aisnet.org/amcis2017/SemanticsIS/Presentations/5/

[38] Donald A. Norman. 2005. *Emotional design: Why we love (or hate) everyday things.* Basic books. https://books.google.co.uk/books?hl=en&lr=&id=h_wAbnGlOC4C&oi=fnd&pg=PA3&dq=emotional+design+&ots=efVGQfn5m9&sig=LCB3ieNHbdf1X_Wf4egX_3fa5WY

[39] Sabine Payr. 2013. *Virtual Butlers and Real People: Styles and Practices in Long-Term Use of a Companion.* Springer. http://link.springer.com/content/pdf/10.1007/978-3-642-37346-6.pdf#page=146

[40] Martin Porcheron, Joel E. Fischer, Moira McGregor, Barry Brown, Ewa Luger, Heloisa Candello, and Kenton O'Hara. 2017. Talking with Conversational Agents in Collaborative Action *(CSCW '17 Companion)*. ACM, New York, NY, USA, 431–436. https://doi.org/10.1145/3022198.3022666

[41] Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied conversational agents in clinical psychology: A scoping review. *Journal of medical Internet research* 19, 5 (2017).

[42] Amanda Purington, Jessie G. Taft, Shruti Sannon, Natalya N. Bazarova, and Samuel Hardman Taylor. 2017. Alexa is my new BFF: Social Roles, User Satisfaction, and Personification of the Amazon Echo. ACM, 2853–2859. http://dl.acm.org/citation.cfm?id=3053246

[43] Rajmohan Rammohan, Nisha Dhanabalsamy, Ves Dimov, and Frank J. Eidelman. 2017. Smartphone Conversational Agents (Apple Siri, Google, Windows Cortana) and Questions about Allergy and Asthma Emergencies. *Journal of Allergy and Clinical Immunology* 139, 2 (2017), AB250.

[44] William E. Scott Jr. 1966. Activation theory and task design. *Organizational behavior and human performance* 1, 1 (1966), 3–30.

[45] Joel Sebastian and Deborah Richards. 2017. Changing stigmatizing attitudes to mental health via education and contact with embodied conversational agents. *Computers in Human Behavior* 73 (2017), 479–488.

[46] Kennon M. Sheldon, Andrew J. Elliot, Youngmee Kim, and Tim Kasser. 2001. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology* 80, 2 (2001), 325. http://psycnet.apa.org/journals/psp/80/2/325/

[47] Alexandre N. Tuch, Rune Trusell, and Kasper Hornbæk. 2013. Analyzing users' narratives to understand experience with interactive products. ACM, 2079–2088. http://dl.acm.org/citation.cfm?id=2481285

[48] Bert Vandenberghe. 2017. Bot Personas as Off-The-Shelf Users. ACM, 782–789. http://dl.acm.org/citation.cfm?id=3052767

[49] Roel Vertegaal, Robert Slagter, Gerrit Van der Veer, and Anton Nijholt. 2001. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. ACM, 301–308.

[50] Alexandra Vtyurina, Denis Savenkov, Eugene Agichtein, and Charles LA Clarke. 2017. Exploring Conversational Search With Humans, Assistants, and Wizards. ACM, 2187–2193. http://dl.acm.org/citation.cfm?id=3053175

[51] David Watson, Lee A. Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063. http://psycnet.apa.org/psycinfo/1988-31508-001

[52] JungKyoon Yoon, Pieter MA Desmet, and Aadjan van der Helm. 2012. Design for interest: Exploratory study on a distinct positive emotion in human-product interaction. *International Journal of Design* 6, 2 (2012).

[53] Jennifer Zamora. 2017. Rise of the Chatbots: Finding A Place for Artificial Intelligence in India and US *(IUI '17 Companion)*. ACM, New York, NY, USA, 109–112. https://doi.org/10.1145/3030024.3040201