User Perceptions of an Intelligent Personal Assistant's Personality*

The Role of Interaction Context

Irene Lopatovska†
School of Information
Pratt Institute
New York, New York USA
ilopatov@pratt.edu

Elena Korshakova
School of Information
Pratt Institute
New York, New York USA
ekorshak@pratt.edu

Diedre Brown
School of Information
Pratt Institute
New York, New York USA
dbrow207@pratt.edu

Yiqiao Li
School of Information
Pratt Institute
New York, New York
USA
ylix56@pratt.edu

Jie Min
School of Information
Pratt Institute
New York, New York
USA
jmin11@pratt.edu

Amber Pasiak
School of Information
Pratt Institute
New York, New York
USA
apasiak@pratt.edu

Kaige Zheng
School of Information
Pratt Institute
New York, New York
USA
kzheng@pratt.edu

ABSTRACT

This paper reports on an experimental study that examined user perceptions of Amazon Alexa's personality, as well as influences of stressful or non-stressful interactions on personality perceptions. The study relied on the Five-Factor Model and the Stereotype Content Model personality frameworks. An online data collection instrument was designed to give Alexa's users (N=50) stressful and non-stressful tasks followed by questions related to Alexa's personality. The assumption was that stressful and non-stressful tasks would frame a user's preference for Alexa's personality. Quantitative data of the participants' ratings of Alexa's responses, and qualitative comments about their ratings, experiences, and thoughts about Alexa were collected and analyzed using descriptive statistics, tests of association, and thematic content analysis. The findings indicated that while the majority of users appreciate Alexa's personality as a way of making interactions more personal and enjoyable, some users prefer their conversational agent to be efficient, robotic-like, and devoid of a personality that might cause attachment. The participants described Alexa as competent in terms of information content and usability, and warm in terms of its personality manifestation, but wanted their ideal Alexa to be even more competent and warm. After experiencing stressful and non-stressful experimental conditions, the participants appreciated Alexa's highly competent and highly warm responses, while disparaging low-warmth/high-competence responses. The participants' comments indicated that after stressful tasks, they were more generous in their assessment of Alexa's performance, and valued manifestations of Alexa's warmth higher than after non-stressful tasks. Additionally, after the second group of tasks/towards the end of the study, the participants tended to score low warmth/high competence responses higher, which might indicate a slight preference for competence after longer interactions or a user's fatigued state. The implications of the findings on the design of conversational interfaces and study limitations are discussed.

CCS CONCEPTS

• Human-centered computing • Human-computer interaction (HCI) • Interaction devices • Sound-based input/output • Information systems • Information retrieval • Users and interactive retrieval

KEYWORDS

Intelligent personal assistant, Conversational agent, Amazon Alexa, Personality, Personification, Anthropomorphizing, Stressful interactions, Conversational interfaces, Branding, Stereotype content model

ACM Reference format:

Irene Lopatovska, Elena Korshakova, Diedre Brown, Yiqiao Li, Jie Min, Amber Pasiak, Kaige Zheng. 2021. User Perceptions of an Intelligent Personal Assistant's Personality: The Role of Interaction Context. In *Proceedings of 2021 of 6th ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR'21), March 14-19, 2021, Canberra, ACT,*

*Article Title Footnote needs to be captured as Title Note

†Author Footnote to be captured as Author Note

CHIIR '21, March 14–19, 2021, Canberra, ACT, Australia

ACM 978-1-4503-8055-3/21/03…$15.00
https://doi.org/10.1145/3406522.3446018

Australia. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3406522.3446018

1 INTRODUCTION

Human tendency to personify objects is a well-known phenomenon [2]. "People respond socially to computers and perceive them as having personalities. Software agents are artifacts that particularly embody those qualities most likely to elicit social responses: fulfilling a social role, using language, and exhibiting contingent behavior [22]." It is particularly easy for humans to attribute personalities to "talking technology", like conversational agents (also known as digital/intelligent/conversational/personal assistants (IPAs)). A conversational agent can be defined as a natural language processing system that supports conversational interactions with users on smart devices, such as cell phones and speakers [9]. It is estimated that approximately 87.7 million U.S. adults were using some sort of conversational device in the first quarter of 2020 [35], with the most popular usually including Amazon Alexa, Google Assistant, and Apple Siri [44]. Manifestations of the IPAs "personality" are evident in the choice of name, voice (male/female), conversational style, and pre-programmed responses related to humor and personality utterances, such as greetings, personal questions and others discussed in the following sections. Amazon Alexa developers even share their aspirations for developing Alexa with an Extraverted/Sensing/Feeling/Judging personality [32]. However, very little is known about users' perceptions of IPA personalities, and whether these perceptions change based on the types of interactions users have with their IPAs. We conducted a study to investigate these questions.

2 RELEVANT LITERATURE

This section provides a definition and major frameworks of personality, outlines the benefits of developing personality in computer agents, and reports on some of the methods of developing personality in computer agents, and specifically, IPAs.

Personality can be defined as a pattern of cognitions, emotions, and behaviors that are relatively stable and consistent over time and across situations [45]. Some of the personality frameworks commonly cited in the human-computer interaction (HCI) literature include:

- The Myers–Briggs Type Indicator (MBTI), grounded in Jungian Typology, is an inventory designed to gauge personal preference on the four dimensions of Extroversion-Introversion (EI), Thinking-Feeling (TF), Sensing–Intuition (SN) and Judgment–Perception (JP) [48].
- The Five Factor model (Big Five or FFM) is based on five dimensions of extraversion, neuroticism, agreeableness, conscientiousness, and openness to experience [45]. While this model correlates to the MBTI with respect to personality dimensions, the Five Factor model provides a broader conceptual framework of personality measurement [46].
- The Multi-source Assessment of Personality Pathology (MAPP) [50] and the Edwards Personal Preference Schedule (EPPS [23]) are self-reported inventories aimed at gauging psychological disorders and are used primarily in clinical psychology.
- The stereotype content model (SCM), is used to assess affective reactions toward objects or interpersonal perceptions, rather than providing personality portraits [26]. This model relies on the dimensions of warmth, perceptions related to friendliness and helpfulness, and competence, perception of skill, ability or intelligence [26].

The benefits of developing products that could be perceived on personality dimensions have long been discussed in marketing literature. For example, the products that are perceived to be warm and competent (based on the SCM framework) are associated with stronger consumer loyalty and engagement [1, 28, 34], increased brand trust [41], positive attitudes [31], and enhanced brand reputation [2]. The benefits of anthropomorphic computer agents are also discussed in the HCI literature. Agents with a personality are shown to improve users' sense of efficacy - one's perceived competence during an interaction with an agent, perceived usefulness of an agent, and facilitate stronger social bonds between people and computers [25]. Bartneck et al. [6] demonstrated that in the context of collaborative interactions, anthropomorphic robot partners were praised more and punished less than the computer, human and machinelike robot partners. According to Gong [29], anthropomorphic agents also enhance the perception of competency and trustworthiness of the presented information, as well as the social responses from users. Recently, Kulms and Kopp [38] found that in an experimental puzzle game with a computer agent, the agent that was perceived as warm increased users' trust and was given more control to assist participants in solving the game. Greater personification has been linked to more sociable interactions with an agent and increased user satisfaction [52]. Personality qualities, like the use of humor or encouraging phrases, have been shown to mitigate performance deficiencies, improve user experiences, emotions and perceptions of systems [4, 36].

A number of studies focus on recommendations for designing system personalities. Some of the earlier work by Belkin [8] examined human to human dialogues and argued that human to computer interactions should simulate the natural language between humans. Although more focused on the elicitation of knowledge and system modelling, the study showed the importance of replicating aspects of human to human interaction in earlier information retrieval research. Recent HCI research in the conversational style of conversational agents similarly suggests that the benefits of a more linguistically aligned language by both the user and the agent can lead to perceptions of trust, reciprocity, and increased engagement by the user with the agent. [12, 39]. In the context of IPAs, Zhou et al. [64] propose three methods for creating an agent's personality through: 1) prosodic factors, such as speaking rate and pitch, 2) physical appearance and language, and 3) interaction behavior with users, such as mannerisms. Goslin et al. [60] findings support the first recommendation of the Zhou et al. [64] framework and show that emotional expressivity through vocal channels, such as a smiling voice, elicits higher trusting behaviors and increases the likeability of an agent, even when evidence of untrustworthiness is present [60]. Another way to design personality is by naming a system. Knott and Kortum [37] found that naming a system leads to more interactions, more complete statements in response to the system prompts, and fewer follow-up prompts, all of which point to better overall quality of interactions.

Companies share very little about their process of designing IPA personalities. Amazon claims that Alexa is designed with the Extraverted/Sensing/Feeling/Judging personality (based on the MBTI classification [48]), and describes Alexa as "approachable, efficient, trustworthy, and natural" [5]. While we do not know much about the intended design of IPA personalities, or how users perceive them, research has produced some recommendations based on user preferences for personalities. In an experimental study of introverted and extroverted conversational agents, Yuting et al. [17] reported user preference for extroverted expressions of personality over introverted types. Zhou et al. [64] tested warm and cheerful, and assertive and serious conversational agents designed to support job interviews. The authors found that job applicants were more willing to confide in and listen to an agent that manifested more assertive and rational personality in high-stakes situations such as job interviews [64]. The situational preferences for personality are also supported by human-to-human interaction research [3, 34, 51], and, indirectly, a large body of personalization research in information seeking/retrieval context [40, 42]. If users' search behaviors and judgements are influenced by context and tasks, so too should their perceptions of systems and their associated personalities.

In summary, prior research suggests that a) humans have a tendency to personify/attribute human qualities to objects; b) personality manifestations (and objects' personification) have benefits in the context of human interactions with systems, and specifically, IPAs; c) general preferences for system's personalities are known (e.g., preferences for warm, supportive, extraverted agents), though these preferences are context-specific; d) there is no clear evidence about how IPA personalities are designed, what users think about them, and whether current IPA personalities are adjustable/situational.

3 METHOD

We designed an experimental study to examine user perceptions of IPA personality and focused specifically on the following questions:

RQ 1. What are users' perceptions of IPA personality, including, 1) what the users' general views on IPAs personality coming into the study, 2) how do the users view the experimental Alexa post-study, 3) what personality traits do users identify as necessary in their ideal Alexa?

RQ 2. Is change in user perceptions of IPA personality dependent on prior situational interactions (stressful/non-stressful)?

The first research question was designed to collect and compare the participants' reactions/thoughts on the existing, experienced and aspirational personality of Alexa. The second question was grounded in the research on the situational preference for warm-competent dimensions of personality [19], as well as the large body of research on the effects of tasks on search behavior. The subsections below describe the key constructs of the study.

3.1 IPA

The Amazon Alexa IPA was used in the study. It is one of the more popular IPAs on the market, and it offers an open user-friendly developer platform compared to other IPAs. All task scenarios and personality-related questions were incorporated into the skill (a voice-activated Alexa app). The skill was developed using the Amazon developer console and VoiceFlow platform. Node.js and JSON syntax were used to program the predetermined interaction model for the tasks and personality questions to have full control over the testing process. The skill was hosted on cloud-based services and was publicly available to all Amazon Alexa users. To launch the experimental skill, existing Alexa users had to activate the skill using the invocation name ("Alexa, open Digital Assistant Research"). New users needed to install the Alexa app, create an Amazon account, log in and activate the Digital Assistant Research skill by saying its invocation name. To ensure that participants were receiving responses from the test skill and not generic Alexa, all the test responses started with the prompt "My response is…". The participants could repeat their utterances if they needed more time to capture Alexa's responses. The use of the experimental skill was guided by the online questionnaire-instruction form described below.

3.2 Personality Type

The study relied on two personality frameworks: the FFM and the SCM. Due to the recent critique of MBTI and its inconveniently long form, FFM has become the preferred model in recent years [10, 58], and we adopted to use it over MBTI. Table A1[1] illustrates the five dimensions of the FFM that were used to determine the dominant traits in user perceptions of the experimental Alexa personality. The FFM has been used in IPA research: Chen et al. [17] used the intra-/extra-version dimensions to design conversational agents and found a preference for extroversion; Castillo et al. [16] proposed using neuroticism and extroversion dimensions of the FFM for designing and evaluating conversational agents; Volkel et. al [59] analyzed personality descriptors of conversational agents using

[1] For Appendix Table A1: https://irenelopatovska.files.wordpress.com/2021/01/appendix_tables_lopatovska_et_al_2021-upofanipap_2.pdf

the FFM, and concluded that the dimensions of artificiality and serviceability should be added to the FFM in the IPA context. We integrated a shorter version of the FFM, the standard 10 Item Big-Five Inventory (BFI-10) [without artificiality and serviceability dimensions since they have not been validated in prior research], into our data collection instrument. The BFI-10 uses two questions to assess each dimension, with a total of 10 items in the inventory [47, 53]. Based on the published recommendations, we added an item to the agreeableness dimension, resulting in a total of 11 item inventory [53].

The SCM framework was used to assess users' immediate reactions to Alexa responses to the personality utterances, as well as to inform a coding schema for participants' comments. The model has been used in the marketing and branding industry, where products are recommended to evoke high warmth and high competence perceptions [1, 30, 34]. A study of embodied virtual agents showed that perceived warmth and competence of an agent increased its perceived believability and, ultimately, user satisfaction [21, 63]. More recently, Biancardi et al. [11] found that the perceived warmth of a conversational agent positively influenced users' satisfaction with interaction and agent's likeability.

3.3 IPA Personality Utterances/Responses

In order to answer RQ2 and understand if the perceptions of Alexa personality are framed by prior immediate interaction types, we needed to embed utterances that would reveal Alexa's personality into the data collection instrument. The utterances related to Alexa's personality that were shown to solicit high-warmth/high-competence (HW/HC), high warmth/low-competence (HW/LC), and low-warmth/high-competence (LW/HC) responses from IPAs were pulled from the classification of IPA utterances/responses [43]. We did not include LW/LC responses as they are usually treated as least desirable, and least liked by users in any situation. For example, Olivera et al. [49] recommend designing for the manifestation where at least one dimension is high in order to keep users engaged and ensure that the interaction between user and agent moves forward. An initial list of 30 utterances, 10 in each category that were previously shown to solicit HW/HC, HW/LC, LW/HC responses from Alexa, were given to 14 participants who rated them as warm and competent on a 7-point Likert scale. Two Alexa responses with the representative scores in the HW/HC, HW/LC, LW/HC, 6 utterances/responses in total were chosen for the test. Examples of the personality utterances and Alexa responses in each warmth-competence dimension are listed in Table 1.

Participants were given different sets of three personality questions after stressful and non-stressful tasks (in order to address RQ2), and asked to provide general ratings of experimental Alexa's responses on a 7-point Likert scale (poor-excellent), rate responses on the warmth/competence dimensions, as well as provide brief comments to explain their ratings. Since in professional and individualistic activities, people show tendency to value competence over warmth [19, 64], our initial assumption was that after stressful tasks, users might value a competent Alexa personality over warm (e.g., if a person does not enjoy the content/complexity of interaction and feels time pressure, they might care more about the informational value of a response (competence), than the niceties and friendliness of a response (warmth).

Table 1: Examples of Alexa Responses to Personality-Testing Questions that Solicited High-Warmth/High-Competence, High-Warmth/Low-Competence, and Low-Warmth/High-Competence Responses

	User Utterance	**Alexa Response**
HW/HC *Friendly and informa-tional*	What should I be for Halloween?	You'll look great in any costume, but I think emoji costumes are fun. Dress in yellow and use paper plates... It's simple.
HW/LC *Friendly, but not to the point*	Am I a good person?	Well, I like you
LW/HC *Informa-tional, but not friendly*	What should I wear today?	The weather forecast is mostly sunny, with a high near 71. Northwest wind from 5 to 10 miles per hour.

3.4 Interaction Types/Scenarios

In order to test the situational dependence of IPA personality preference, we designed two types of interactions: stressful and non-stressful/enjoyable. Stressful interaction was simulated by requiring participants to ask Alexa four questions on unpleasant or work-like topics, receive complex informational response and record their answers in under 2 minutes. Though the time limit was not enforced, it was used to create additional discomfort for the participants. The four "stressful" utterances were inspired by prior classification of tasks in the information retrieval context [7, 15, 54, 55, 57], and included information requests about:

- The route to the closest hospital for a neighbor who has a heart attack
- Finding a tax bracket and tax rate information
- What to do after coming into contact with someone positive for COVID-19, and
- Completing/verifying business-related information for a work email.

The four utterances used to simulate "enjoyable" interactions were largely informed by leisure information behavior research/task scenarios [18, 24, 62], and included:

- A request for interesting information about Bora-Bora for a vacationing friend
- A test of a stress-management feature (where participants were asked to listen and rate a short meditation instruction)
- A request for a joke, and
- A new movie recommendation.

The full text of stressful and non-stressful utterances and test Alexa's responses can be found in Appendix Table A2[2].

The final utterances for the study were selected from a longer list of stressful and non-stressful utterances that were pretested by 14 participants during a pilot study. The final study participants also had a chance to rate the utterances as "stressful" and "fun" and confirmed our classification (with hospital route and COVID-19 scenarios being rated as the most stressful, and Bora-Bora and the joke request rated as the most fun by the majority of the participants). All of the utterances and responses were designed to simulate real interactions, and the real Alexa responses informed the preprogrammed responses of the experimental skill to as great of an extent as possible.

The participants were asked to rate Alexa's response to each stressful or non-stressful utterance using the 7-point Likert scale and provide an explanation for their rating through the required comment field on the questionnaire form. Due to the constraints of this article's length, and the fact that the participants' performance on (non)stressful tasks were peripheral to the main RQs, task performance findings are not reported here.

3.5 Experimental Setting

The study was initially planned as a laboratory experiment, where participants would interact with the test Amazon Alexa skill on a standalone Echo device. The participants would follow an experimental protocol to go through interactions and assess/comment on them in a questionnaire, as well as have an interview with a researcher. However, due to the COVID-19 lockdown, the experiment had to be adjusted to a fully remote format. A questionnaire was developed to guide the participants through the study and collect data. The questionnaire included items related to:

- The participants' demographics and their IPA usage/ownership/experience
- The participants' general views on IPAs personality
- Instructions for downloading the Alexa app/activating the experimental skill on a mobile device (so that the questionnaire could be filled on the other device)
- The experimental tasks/scenarios: stressful and non-stressful tasks, the order of which was alternated based on the participant's month of birth (even/odd) to avoid order effect
- Ratings of experimental tasks with comments
- Post-task questions about the Alexa personality
- Ratings of Alexa personality on Warm/Competence dimensions, with comments
- Ratings of the experimental task as fun/stressful
- General comments about Alexa in the study and the participant's ideal Alexa.

[2] For Appendix Table A2, refer to: https://irenelopatovska.files.wordpress.com/2021/01/appendix_tables_lopatovska_et_al_2021-upofanipap_2.pdf

The whole test lasted about 40 minutes. The participants received monetary compensation for their participation. The study was approved by the IRB.

3.6 Sample

A total of 50 participants was recruited from the Pratt Institute School of Information listserv. The participants were equally split into two groups: a group that was asked to perform stressful tasks first, and a group that performed non-stressful tasks first. Female participants comprised 64% (*N*=32) of the total sample. Sixty-four percent of participants were 25-34 y.o., with 18-24 y.o. being the second largest group (*N*=11/22%). The primary language of the 74% of participants was English, (*N*=37), with Chinese (*N*=5/10%), Korean (*N*=3/6%), German (*N*=1/2%), and Polish (*N*=1/2%) also being represented in the sample (we asked about primary language to gauge the cultural background of the participants). Fifty-six percent of participants (*N*=28) considered themselves to be active IPA users, and 80% of these participants used a dedicated device(s) for their IPAs. The majority of the active users were long term users of Apple Siri. The IPAs were most frequently used to play music (*N*=24), set an alarm/timer (*N*=24), or to check the weather (*N*=20). Of the participants who reported not using IPAs (*N*=22/44%), ten felt that they did not need them and four indicated distrusting the IPAs or having privacy concerns.

4 FINDINGS

4.1 IPA Personality Perceptions

At the beginning of the study, we asked the participants if they think an IPA should have a personality. Sixty-one percent of the participants responded that IPAs should have human-like features and human personalities. This was further supported by the participants' comments which stated that a human-like voice and personality make interactions more entertaining and enjoyable, and support emotional or personalized experiences with their IPAs. The participants who did not believe that an IPA should have personality think that personalities get in the way of their IPA completing tasks, make IPAs "creepy" and "scary", are not necessary in a robotic assistant or they do not wish to develop a relationship with their device in the same manner as they do with another human.

The FFM inventory of the experimental Alexa's personality completed by participants right after the experimental tasks indicated that users perceived it as Extraverted, Highly Agreeable/Sympathetic, Highly Conscious/Organized, Low Neurotic/Resilient, and Highly Open/Open minded. While this paper does not focus on the relationship between participants' personalities/demographics and their judgements of Alexa, we will briefly report that we found medium size association between FFM scores of participant personalities and FFM scores they assigned to Alexa. We found positive correlations in the dimensions of conscientiousness ($r(48) = 0.28$, $p > .001$), openness (($r(48) = 0.21$, $p > .001$), agreeableness ($r(48) = 0.18$, $p > .001$), and extraversion (($r(48) = 0.12$, $p > .001$), and negatively correlation on the neuroticism dimension ($r(48) = -.18$, $p > .001$) (participants' FFM totals are reported in Table 2).

The participants' comments about Alexa were analyzed using qualitative thematic approach [13] aimed at "identifying, analyzing and reporting patterns within data." Three independent researchers examined the comments data to identify common themes, which represent patterns of meaning that were mentioned repeatedly. Initial themes were organized into a coding schema (an example of the coding schema is provided in Appendix Table A3[3]). Out of a total of 149 descriptors of Alexa, the participants referenced interface usability (which included communication quality) most frequently (40%), followed by IPA personality (35%), information/content quality (23%), and affective reactions (1%) (Table 3).

Table 2: Participants' Personality and Their Perception of the Test Alexa Personality based on FFM Inventory

Dimensions	Description	Alexa (N/%)	Participants (N/%)
Extraversion	Extrovert	27/54%	27/54%
	Introvert	23/46%	23/46%
Agreeable-ness	Sympathetic	48/96%	49/98%
	Antagonistic	2/4%	1/2%
Conscientious-ness	Organized	43/86%	43/86%
	Disorganized	7/14%	7/14%
Neuroticism	Sensitive	18/36%	29/58%
	Resilient	32/64%	21/42%
Openness	Open-Minded	37/74%	49/98%
	Conservative	13/26%	1/2%

Table 3: Themes in Comments about Alexa in the Study and Ideal Alexa

Theme ***Examples of Comments***	**Alexa in this study (N/%)**	**Ideal Alexa (N/%)**	**Difference (%)**
IPA personality *Friendly, Funny*	52/34.9	58/40.0	+5.1
Interface Usability *Effective, Error free*	60/40.3	41/28.3	-12
Information content/quality *Informative*	35/23.5	45/31	+7.5
Affective experience of user *Satisfactory*	2/1.3	1/0.7	-0.6

The patterns of descriptions of an ideal Alexa were largely similar to the descriptors of the current Alexa. Out of a total 145 descriptors, comments about Alexa's personality were most frequent (40%), followed by comments about content quality (31%), interface usability (28%), and affective reactions (single response). Table 3 summarizes the differences between the users' descriptions of the current Alexa and their ideal Alexa.

We also analyzed how Alexa descriptors mapped onto the Warmth/Competence framework. The participant responses were split relatively evenly between adjectives referencing competence (47%) and those referencing warmth (53%). While most of the competence-related responses acknowledged Alexa's high-competence (40% of overall results, 83% of competence-related results), some responses did address the low perceived competence of Alexa (8% of all comments, 17% of competence responses), indicating that users were not entirely satisfied with the information content provided by Alexa in the study. The results between high- and low-warmth adjectives were evenly split (26% of overall results each, 50% of warmth-related results each, Table 4).

Table 4: Comments about Warmth and Competence of Alexa in the Study Alexa and Ideal Alexa

Comments about Alexa's Competence/ Warmth ***Examples of Comments***	**Test Alexa in the study (N/%)**	**Ideal Alexa (N/%)**	**Difference (%)**
Competence: High *Accurate, Competent*	59/39.6	77/53.1	+13.5
Competence: Low *Non-functional, Limited*	12/8	1/0.7	-7.4
Competence: Total	**71/47.6**	**78/53.8**	**+6.1**
Warmth: High *Empathetic, Kind*	39/26.2	53/36.5	+10.4
Warmth: Low *Insincere, Boring*	39/26.2	14/9.7	-16.5
Warmth: Total	**78/52.3**	**67/46.2**	**-6.1**
Overall Total	**149/100**	**145/100**	

The relatively even distribution of the comments between warmth- and competence-related adjectives were reflected in the users' descriptions of their ideal Alexa, as shown in Table 3. We noted a 6% shift from more comments about the warmth of the study's Alexa, to 6% more comments about the competence of an ideal Alexa.

4.2 Situational Personality Perceptions

We examined whether the participants' ratings of Alexa's responses to personality questions varied after stressful and non-stressful tasks. The mean scores of the participants' responses indicate a stronger preference for HW/HC responses, especially after stressful tasks (whether the task was first or following a non-stressful task). The HW/LC responses also received above average ratings across tasks and rotations (Table 5). However, the ratings of the LW/HC responses differed between the first and second groups of tasks. This result was further confirmed by

[3] For Appendix Table A3, refer to: https://irenelopatovska.files.wordpress.com/2021/01/appendix_tables_lopatovska_et_al_2021-upofanipap_2.pdf

an ANOVA test ($F(1,98) = .15$, $p < .001$) that pointed to the only statistically significant difference in mean scores for LW/HC responses: the average LW/HC score after the second/final tasks was significantly higher ($M = 5.2$) then after the first tasks ($M = 3.8$), regardless of whether these tasks were stressful or non-stressful.

The content analysis of the participants' justifications for their ratings of Alexa responses on warmth-competence dimensions (Table 6) showed that most frequent comments focused on the content quality of Alexa's response, followed by comments about its personality and interface usability.

Content quality is an indirect manifestation of competence dimension, and the high frequency of comments related to it illustrates that competence is important in the overall judgement of an IPA's response. The high frequency of personality related comments suggests that manifestations of warmth and friendliness were also important to the users. It is worth noting that after the stressful tasks, the number of positive comments related to Alexa's personality was higher. This may point to a higher appreciation of warm responses after stressful tasks. After the non-stressful tasks, participants were more critical of manifestations of Alexa's personality, its usability and communication styles, as the number of negative comments in these categories increased.

5 DISCUSSION

5.1 IPA Personality Perceptions

The majority of our participants outlined the benefits of an IPA personality in creating human-like, personalized conversational interactions. However, some participants were not convinced of the merits of an IPA personality, and cited their preferences for more "utilitarian", transactional interactions, as well as fears of having or developing relationships with IPAs in their rationales. While the research in multiple disciplines usually outlines the inevitability and the benefits of personification, the fact that not all users want their IPAs to have personas highlights the need to develop a customizable interface with an optional persona for IPAs like Alexa.

Our participants identified the personality of the experimental Alexa as Extraverted, Highly Agreeable/Sympathetic, Highly Conscious/Organized, Low Neurotic/Resilient, and Highly Open/Open minded, a perception that aligns with Alexa developers' intentions to design an Extraverted/Sensing/Feeling/Judging (ESFJ) Alexa [32] (though we used the FFM and Amazon is citing the MBTI frameworks, there is a large overlap between the personality dimensions of the two models [46]). The alignment of user perceptions of the experimental Alexa skill and intended Alexa personality might have two explanations: a) our experimental skill with its embedded interactions managed to closely capture the "real" ESFJ Alexa personality designed by Amazon, or b) our skill came close to the ideal, inspirational Alexa that might be far from reality. A separate study is needed to investigate how Alexa users perceive its personality in a naturalistic setting and non-

Table 5: Mean Ratings of the Participants' Overall Satisfaction with Alexa's Responses to Personality Questions After Each Task Based on Warmth and Competence

	Overall Rating After			
	1st Task (M (SD))		2nd Task (M (SD))	
	Stressful Group 1	Non-Stressful Group 2	Non-Stressful Group 1	Stressful Group 2
HW/HC	5.2 (1.4)	4.7 (1.6)	5.8 (1.3)	4.6 (2)
HW/LC	5.2 (1.5)	4.3 (1.6)	4.6 (1.6)	4.1 (1.9)
LW/HC	3.8 (2)	4 (2)	5.2 (1.5)	5.4 (1.5)

Table 6: Main Themes in Participants' Justifications of Rating of Alexa's Response to Personality Questions after Stressful and Non-Stressful Tasks

Themes in Comments ***Examples of Comments***	**Positive (N/%)**	**Negative (N/%)**	**Neutral (N/%)**	**Total (N/%)**
After Stressful Tasks				
Information/Content Quality *Information was helpful to me*	60/56	26/24	21/20	107/39
IPA Personality *I like the sense of humor*	63/77	15/18	4/5	82/30
Interface Usability *Redirected me to the website*	15/30	34/68	1/2	50/18
Affective User Reaction *I was disappointed*	22/76	6/21	1/3	29/11
Unclear				8/3
Total	**160/58**	**81/29**	**27/10**	**276/100**
After Non-Stressful Tasks				
Information/Content Quality *Response was very accurate*	68/52	38/29	24/18	130/41
IPA Personality *Like a machine too much without emotions*	56/63	31/35	2/2	89/28
Interface Usability *Prefer visual*	8/17	37/77	3/6	48/15
Affective User Reaction *I feel satisfied*	27/60	18/40	0/0	45/14
Unclear				4/1
Total	**159/50**	**124/39**	**29/9**	**316/100**

experimental interactions. Our analysis of the participants' comments indicated that they generally assess the Alexa as competent and warm, but would like their ideal Alexa to be even more competent (in terms of content quality and usability) and warm (usability/communication-emotional expressivity). The findings that users want their Alexa to score high on both competence and warmth personality dimensions confirms general guidelines for developing popular brands [1]. The fact that the SCM and its warmth-competence dimensions are reflected in the users' comments about Alexa opens up a potential avenue for research around the following questions: what other information systems are perceived to have personalities and/or manifestations of warmth-competence? For example, do users perceive Google's search engine to have a personality, to express "warmth", and not just provide informational value, "competence"? Could user expectations for warmth and competence be used to classify information systems into the ones where users primarily expect "competence" (information), "warmth" (emotional support, casual conversation), or both (which might be the case of IPAs)? A fact that users engage in interactions that aim to test an IPA's personality [42] attest to its differences compared to the more traditional information retrieval systems.

5.2 Interaction-dependent Personality Perceptions

Overall, we did not find significant differences in the participants' appreciation for warm/competent personality responses after stressful and non-stressful tasks. After both types of tasks, participants rated HW/HC responses highly. The participants also gave consistently high scores to HW/LC responses, which might indicate the importance of warmth over competence in casual interactions aimed at understanding Alexa's personality. A statistically significant change in participants' ratings of personality responses occurred with LW/HC personality responses. The rating of low-warmth responses was not affected by the stressful/non-stressful condition, but it was affected by the sequence and timing of LW/HC responses: after the first set of tasks, the ratings of LW/HC responses were significantly lower than ratings of the same type of responses after the second and final set of tasks. This result may suggest that a preference for competence over warmth might be less dependent on the type of prior tasks, and more on user fatigue. It is possible that as the experiment progressed, the participants wanted it to end quickly and appreciated the informational aspects of responses that would aid them in completing the study faster. Another possible explanation is the longer interactions with Alexa (end of second tasks) led to a higher appreciation of its competence, compared to warmth. This hypothesis is supported by prior literature that suggests the primacy of warmth in initial judgements of people and objects, but as the initial importance of warmth fades away, people are more likely to appreciate competence [26]. These findings suggest that while it is important to design "warm" personalities to attract users and create first impressions, developing competence perceptions might be more important for longer interactions and/or long-term IPA use [20, 56].

While the quantitative data (response ratings) did not reveal major differences in personality response ratings after stressful and non-stressful tasks, the qualitative data of the participants' comments tells a richer story. After stressful tasks, the users were less critical of Alexa's responses and offered more positive comments pertaining to the manifestations of personality. After non-stressful tasks, the participants offered more comments overall, including more comments about the content quality and their affective reactions. The participants' comments after stressful and non-stressful tasks might indicate that manifestations of IPA personality are better appreciated after stressful tasks, when users' are seeking emotional support after a stressful experience. It is also possible that after experiencing the stressful task, the participants may have rationalized their experience to three "essential aspects of the situation" [33]: stressful event—Alexa provided competent information—stressful event completed. Here, and as supporting prior anthropomorphic studies in HCI state, the participant may have viewed Alexa as a collaborator who helps complete the task. While it is unclear why after non-stressful tasks the users were more critical and verbose, it might be related to not having time pressures translating into spending more time thinking and critiquing Alexa responses. When we designed the study, we had a different assumption that the "negative" frame of stressful tasks will "spill over" on the assessments of Alexa's personality [27, 61], and will make users more critical of Alexa. While our hypothesis did not confirm, we think it is worth continuing to investigate situational effects like stress on users' preferences for IPA's personality.

6 CONCLUSION

5.1 IPA Personality Perceptions

We conducted an experiment to understand users' perceptions of personality in IPAs, specifically, Amazon Alexa. The study also examined if personality perceptions were situation-dependent and framed by the preceding stressful and non-stressful interactions. The findings relied on the qualitative data of the participants' ratings of Alexa's responses, as well as the qualitative comments about their ratings, experiences and thoughts about generic IPAs and the experimental Alexa.

The findings suggest that not all users want their IPAs to have personalities, but those who do, see its merits in making interactions more entertaining, enjoyable, and personal. The participants classified experimental Alexa as having Extraverted, Highly Agreeable/Sympathetic, Highly Conscious/Organized, Low Neurotic/Resilient, and Highly Open/Open minded personality, which aligns with Alexa developers' intentions [32], but might not represent user interactions with the "real" Alexa. The participants described the experimental Alexa as competent in terms of the provided information content, and warm in terms of its presentation styles and personality manifestation, but wanted even more competence and warmth in their ideal Alexa.

Despite going through stressful and non-stressful experimental conditions, participants always appreciated highly competent and highly warm responses aimed at gauging Alexa personality. While the numeric scores of Alexa responses to personality questions did not show a statistically significant difference (for the most part), participants comments indicated that after stressful tasks, the participants were more generous in their assessment of Alexa's performance and tended to value manifestations of Alexa's warmth higher than after non-stressful tasks. We also found that after the second tasks/towards the end of the study, participants tended to score low warmth/high competence responses higher, indicating a slight preference for competence after longer interactions or in a more fatigued state. The findings lead us to make the following recommendations:

- Since not all IPA users see the benefits of their IPAs having personalities, and would prefer interfacing with a robotic efficient machine, an option of disabling any personality manifestations should be offered to its users.
- Since IPA users tend to value manifestations of high warmth and high competence (similar to the consumers of other products/brands), IPA designers should aim to calibrate IPA's responses toward the "golden quadrant" of high warmth and high competence [1, 49].
- Since the length and the type of interaction might affect users' perceptions and preferences for warmth and/or competent manifestations of IPA personalities, the IPA responses should consider interaction context and duration in designing responses that lean more towards friendly/warm or efficient/competent dimensions.

An additional recommendation came out of the experimental design. While we relied heavily on earlier information retrieval (IR) studies in simulating stressful and non-stressful tasks, we identified a gap in retrieval/interaction tasks classifications. We would recommend expending classifications of user interactions in IR and/or conversational systems to include enjoyable/non-stressful and unpleasant/stressful and exploring their effects on subsequent user behaviors.

The study had a number of limitations. Due to sample size, some of the statistical tests might not have shown significant results, even when qualitative data pointed to such differences in user reactions to Alexa manifestations after stressful and non-stressful tasks. Alexa's manifestations of personality were primarily operationalized through six utterances that were previously shown to solicit HW/HC, HW/LC, and LW/HC personality perceptions from its users. The decision to only use six "personality" utterances was guided by the considerations of experiment's length and control over stimuli/utterances. However, it might have resulted in unrepresentative judgments of Alexa's warmth and competence. Another limitation is the reliance on a relatively homogeneous sample of largely graduate students [younger, more tech savvy than other groups, primarily English speaking]. As our questionnaire collected data on the participants' demographics and familiarity with Alexa, we believe that our sample of participants was to some degree representative of a broader population of IPA users and non-users. However, we would encourage more research with IPA users and non-users with different cultural and linguistic backgrounds as these characteristics have been show to affect personality preferences and perceptions. The transition to a fully remote format might have introduced additional challenges. As the researchers were not present to provide instruction or conduct post-experiment interviews with the participant, it was not fully possible to ensure that the participants were following the experiment instructions to the letter. However, as the study was pretested with 14 participants, the researchers were equipped to identify and fix most of the issues that emerged by following the fully online asynchronous experimental procedure. Additionally, as the participants' mental stress experienced by having to alter their behavioral patterns due to the COVID-19 lockdown could not be assessed, unintentional responses to stress could have fostered a spillover effect to the study behaviors. The study used the Amazon Alexa platform to examine user interactions with an IPA, which makes findings not widely generalizable to other IPAs. A naturalistic non-experimental study conducted under more normal circumstances, with more culturally/linguistically diverse sample, and with multiple IPAs could mitigate some of these limitations.

Despite the limitations, the study contributed to our understating of user perceptions of IPAs, and proposed helpful frameworks to examine and design IPA personality manifestations. At present, there is no clear evidence as to what IPA personalities are, whether they are adjustable/situational, and what users think about them. In fact, a recent study found that users tend to perceive IPAs as primarily an interface to their phone, web or another system, and only peripherally as a "handy helper" which plays the role of a quick and helpful assistant, and a "repository of knowledge" which is linked to a huge collection of knowledge [13]. If indeed manifestations of personality in IPA are beneficial for its users, we need to better understand how different users perceive them in various contexts, and turn conversational assistants into truly "handy helpers" and companions.

ACKNOWLEDGMENTS

We would like to thank Adrienne Lang, Anna Flinchbaugh, Catherine Poole, Felicia Tu, Leah Carlson-Downie, Michael Lewis, Pei Chun Ou Yang, Peter Defenderfer and Xueping Wu for their contributions to the project. The study was partially supported by the Pratt Institute School of Information Faculty Innovation Fund.

REFERENCES

[1] Jennifer L. Aaker, Emily N. Garbinsky, and Kathleen D. Vohs. 2012. Cultivating admiration in brands: Warmth, competence, and landing in the "golden quadrant." Journal of Consumer Psychology 22, 2, 191–194. DOI: https://doi.org/10.1016/j.jcps.2011.11.012

[2] Jennifer L. Aaker, Kathleen Vohs and Cassie Mogilner. 2010. Non-Profits Are Seen as Warm and For-Profits as Competent: Firm Stereotypes Matter. Journal of Consumer Research 37, 2, 224–237. DOI: https://doi.org/10.2139/ssrn.1540134

[3] Andrea E. Abele and Susanne Brack. 2013. Preference for other persons' traits is dependent on the kind of social relationship. Social Psychology 44, 2 (2013), 84–94. DOI: https://doi.org/10.1027/1864-9335/a000138

[4] Shazia Afzal, Bryan Dempsey, Cassius D'Helon, Nirmal Mukhi, Milena Pribic, Aaron Sickler, Peggy Strong, Mira Vanchiswar, and Lorin Wilde. 2019. The Personality of AI Systems in Education: Experiences with the Watson Tutor, a one-on-one virtual tutoring system. Childhood Education 95, 1, 44–52. DOI: https://doi.org/10.1080/00094056.2019.1565809

[5] Amazon Alexa (2019). Amazon Echo & Alexa Brand Guidelines. https://m.mediaamazon.com/images/G/01/mobile-apps/dex/alexa/alexaauto/EchoAlexaGuideExternal.pdf

[6] Christoph Bartneck, Juliane Reichenbach, and Julie Carpenter. 2006. Use of Praise and Punishment in Human-Robot Collaborative Teams. In ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication, 177-182. DOI: 10.1109/ROMAN.2006.314414

[7] Scott S. Bateman, Carl A. Gutwin, and Gordon I. McCalla. 2013. Social navigation for loosely-coupled information seeking in tightly-knit groups using webwear. In Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13). Association for Computing Machinery, New York, NY, USA, 955–966. DOI:https://doi-org.ezproxy.pratt.edu/10.1145/2441776.244188

[8] Nicholas J. Belkin, Helen M. Brooks, and Penny J. Daniels. 1987. Knowledge elicitation using discourse analysis. International Journal of Man-Machine Studies 27, 2, 127–144. DOI:https://doi.org/10.1016/S0020-7373(87)80047-0

[9] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottrjdge. 2018. Understanding the Long-Term Use of Smart Speaker Assistants. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 3, 1–24. DOI:https://doi.org/10.1145/3264901

[10] Tammy L. Bess and Robert J. Harvey. 2002. Bimodal score distributions and the Myers-Briggs Type Indicator: Fact or artifact? Journal of Personality Assessment 78, 1, 176–186. DOI:https://doi.org/10.1207/S15327752JPA7801_11

[11]] Beatrice Biancardi, Maurizio Mancini, Paul Lerner, and Catherine Pelachaud. 2019. Managing an Agent's Self-Presentational Strategies During an Interaction. Frontiers in Robotics and AI 6. DOI:https://doi.org/10.3389/frobt.2019.00093

[12] Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. McLean. 2010. Linguistic alignment between people and computers. Journal of Pragmatics 42, 9, 2355-2368. DOI:https:// doi.org/10.1016/j.pragma.2009.12.012

[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 2, 77–101. DOI:https://doi.org/10.1191/1478088706qp063oa (p.79)

[14] Raluca Budiu. 2019. Mental Models for Intelligent Assistants. Retrieved June 10, 2020 from https://www.nngroup.com/articles/mental-model-ai-assistants/

[15] Katriina Bystrom and Kalervo Jarvelin. 1995. Task Complexity Affects Information Seeking and Use. Information Processing & Management 31, 2, 191–213.

[16] Linda G. Castillo, Flor V. Perez, Rosalinda Castillo, and Mona R. Ghosheh. 2010. Construction and initial validation of the Marianismo Beliefs Scale. Counselling Psychology Quarterly 23, 2, 163–175. DOI:https://doi.org/10.1080/09515071003776036

[17] Yuting Chen, Adeel Naveed, and Robert Porzel. 2010. Behavior and preference in minimal personality: a study on embodied conversational agents. In International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction (ICMI-MLMI '10). Association for Computing Machinery, New York, NY, USA, Article 49, 1–4. DOI: https://doi.org/10.1145/1891903.1891963

[18] Michael Crabb and Vicki Hanson. 2016. An analysis of age, technology usage, and cognitive characteristics within information retrieval tasks. ACM Transactions on Accessible Computing 8, 3, 10. DOI:https://doi.org/10.1145/2856046

[19] Amy J. C. Cuddy, Susan T. Fiske, and Peter Glick. 2008. Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. In Advances in Experimental Social Psychology. Academic Press, 61–149. DOI:https://doi.org/10.1016/S0065-2601(07)00002-0

[20] Amy J. C. Cuddy, Matthew Kohut, and John Neffinger. 2013. Connect, Then Lead. Harvard Business Review. Retrieved October 1, 2020 from https://hbr.org/2013/07/connect-then-lead

[21] Virginie Demeure, Radosław Niewiadomski, and Catherine Pelachaud. 2011. How Is Believability of a Virtual Agent Related to Warmth, Competence, Personification, and Embodiment? Presence: Teleoperators and Virtual Environments 20, 5, 431–448. DOI:https://doi.org/10.1162/PRES_a_00065

[22] Christopher D. Dryer. 1999. Getting personal with computers: How to design personalities for agents. Applied Artificial Intelligence 13, 3 (April 1999), 273–295. DOI: https://doi.org/10.1080/088395199117423 (page 273)

[23] Allen L. Edwards. 1959. Edwards personal preference schedule. Psychological Corporation, New York, NY.

[24] David Elsweiler, Max L. Wilson, and Brian Kirkegaard Lunn. 2011. Understanding Casual-Leisure Information Behaviour. New Directions in Information Behaviour, 211–241. DOI:https://doi.org/10.1108/S1876-0562(2011)002011a012

[25] Nicholas Epley, Adam Waytz, and John T. Cacioppo. 2007. On seeing human: A three-factor theory of anthropomorphism. Psychological Review 114, 4, 864–886. DOI: https://doi.org/10.1037/0033-295X.114.4.864

[26] Susan T. Fiske, Amy J. C. Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: warmth and competence. Trends in Cognitive Sciences 11, 2, 77–83. DOI: https://doi.org/10.1016/j.tics.2006.11.005

[27] Joseph Forgas. 1995. Mood and judgment: The Affect Infusion Model (AIM). Psychological bulletin 117, 39–66. DOI:https://doi.org/10.1037/0033-2909.117.1.39

[28] Susan Fournier and Claudio Alvarez. 2012. Brands as relationship partners: Warmth, competence, and in-between. Journal of Consumer Psychology 22, 2, 177–185. DOI: https://doi.org/10.1016/j.jcps.2011.10.003

[29] Li Gong. 2008. How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. Computers in Human Behavior 24, 4, 1494–1509. DOI: https://doi.org/10.1016/j.chb.2007.05.007

[30] Georgios Halkias and Adamantios Diamantopoulos. 2020. Universal dimensions of individuals' perception: Revisiting the operationalization of warmth and competence with a mixed-method approach. International Journal of Research in Marketing. DOI:https://doi.org/10.1016/j.ijresmar.2020.02.004

[31] Bjoern S. Ivens, Alexander Leischnig, Brigitte Muller, and Katharina Valta. 2015. On the Role of Brand Stereotypes in Shaping Consumer Response toward Brands: An Empirical Examination of Direct and Mediating Effects of Warmth and Competence. Psychology & Marketing 32, 8, 808–820. DOI: https://doi.org/10.1002/mar.20820

[32] Khari Johnson. 2018. Amazon says Alexa is an ESFJ (Updated). Retrieved June 27, 2020 from https://venturebeat.com/2018/10/25/amazon-says-alexa-is-an-estj/

[33] Daniel Kahneman and Shane Frederick. 2007. Frames and brains: Elicitation and control of response tendencies. Trends in Cognitive Sciences, 11(2), 45–46. DOI:https://doi.org/10.1016/j.tics.2006.11.007

[34] Nicolas Kervyn, Susan T. Fiske, and Chris Malone. 2012. Brands as intentional agents framework: How perceived intentions and ability can map brand perception. Journal of Consumer Psychology 22, 2, 166–176. DOI:https://doi.org/10.1016/j.jcps.2011.09.006

[35] Bret Kinsella. 2020. Introducing Voicebot Research - The Most Comprehensive Data on Voice Assistants and AI. Retrieved June 3, 2020 from https://voicebot.ai/2020/04/30/introducing-voicebot-research-the-most-comprehensive-data-on-voice-assistants-and-ai/

[36] Will Knight. 2012. Social intelligence: Siri May Not be the Smartest AI in the World, But it's the Most Socially Adept.Technology Review 115, 80–82.

[37] Benjamin Knott and Phil Kortum. 2006. Personification of Voice User Interfaces: Impacts on User Performance. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting 50, October 2006, 599–603. DOI: https://doi.org/10.1177/154193120605000411

[38] Philipp Kulms and Stefan Kopp. 2018. A Social Cognition Perspective on Human–Computer Trust: The Effect of Perceived Warmth and Competence on Trust in Decision-Making With Computers. Frontiers in Digital Humanities 5, 14. DOI: https://doi.org/10.3389/fdigh.2018.00014

[39] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. International Journal of Human-Computer Studies, 103, 95-105. DOI:https:// doi.org/10.1016/j.ijhcs.2017.02.005

[40] Yuelin Li and Nicholas J. Belkin. 2008. A faceted approach to conceptualizing tasks in information seeking. Information Processing & Management 44, 6, 1822–1837. DOI:https://doi.org/10.1016/j.ipm.2008.07.005

[41] Fuan Li, Nan Zhou, Rajiv Kashyap, and Zhilin Yang. 2008. Brand trust as a second-order factor An alternative measurement model. International Journal of Market Research 50, 817–839. DOI: https://doi.org/10.2501/S147078530820

[42] Jingjing Liu, Chang Liu, and Nicholas J. Belkin. 2020. Personalization in text information retrieval: A survey. Journal of the Association for Information Science and Technology 71, 3, 349–369. DOI:https://doi.org/10.1002/asi.24234

[43] Irene Lopatovska. 2020. Personality Dimensions of Intelligent Personal Assistants. In Proceedings of SIGIR Conference on Human Information Interaction & Retrieval 2020 (CHIIR'20). March 14-18, 2020, Vancouver, British Columbia, Canada, ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3343413.3377993

[44] Gustavo López, Luis Quesada, and Luis Guerrero. 2017. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-based Natural User Interfaces. Advances in Human Factors and Systems Interaction. AHFE 2017. Advances in Intelligent Systems and Computing 592, 241–250. DOI:https://doi.org/10.1007/978-3-319-60366-7_23

[45] Robert R. McCrae and Paul T. Costa. 2003. Personality in Adulthood: A Five-factor Theory Perspective. Guilford Press, New York, NY.

[46] Robert R. McCrae and Paul T. Costa. 1989. Reinterpreting the Myers-Briggs Type Indicator From the Perspective of the Five-Factor Model of Personality.

Journal of Personality 57, 1, 17–40. DOI:https://doi.org/10.1111/j.1467-6494.1989.tb00759.x

[47] Sarah Luisa Müller and Anja Richert. 2018. The Big-Five Personality Dimensions and Attitudes to-wards Robots: A Cross Sectional Study. In Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference (PETRA '18), Association for Computing Machinery, New York, NY, USA, 405–408. DOI:https://doi.org/10.1145/3197768.3203178

[48] Isabel B. Myers, Linda K. Kirby, and Katharine D. Myers. 1998. Introduction to Type: A Guide to Understanding Your Results on the MBTI Instrument (6th ed). CPP, In., Mountain View, CA.

[49] Raquel Oliveira, Patricia Arriaga, Filipa Correia, and Ana Paiva. 2019. The Stereotype Content Model Applied to Human-Robot Interactions in Groups. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE, Daegu, Korea (South), 123–132. DOI:https://doi.org/10.1109/HRI.2019.8673171

[50] Thomas F. Oltmanns and Eric Turkheimer. 2006. Perceptions of Self and Others Regarding Pathological Personality Traits. In Personality and psychopathology, R. F. Krueger & J. L. Tackett, Eds. The Guilford Press, New York, NY, 71–111.

[51] Caitlin M. Porter and James R. Rigby. 2019. Relationship context and personality shape people's preferences for network relationship partners. Personal Relationships 26, 2, 310–330. DOI:https://doi.org/10.1111/pere.12275

[52] 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. In Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17). Association for Computing Machinery, New York, NY, USA, 2853–2859. DOI: https://doi.org/10.1145/3027063.3053246

[53] Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. Journal of Research in Personality 41, 1, 203–212. DOI:https://doi.org/10.1016/j.jrp.2006.02.001

[54] Ian Ruthven and Diane Kelly. 2011. Interactive Information Seeking, Behaviour and Retrieval. Facet Publishing.

[55] Miamaria Saastamoinen and Kalervo Järvelin. 2017. Search task features in work tasks of varying types and complexity. Journal of the Association for Information Science & Technology 68, 5, 1111–1123.

[56] Clemens Stachl, Quay Au, Ramona Schoedel, Samuel D. Gosling, Gabriella M. Harari, Daniel Buschek, Sarah Theres Völkel, Tobias Schuwerk, Michelle Oldemeier, Theresa Ullmann, Heinrich Hussmann, Bernd Bischl, and Markus Bühner. 2020. Predicting personality from patterns of behavior collected with smartphones. PNAS 117, 30, 17680–17687. DOI:https://doi.org/10.1073/pnas.1920484117

[57] Robert Stebbins. 2009. Serious Leisure and Work. Sociology Compass 3, 5, 764–774. DOI:https://doi.org/10.1111/j.1751-9020.2009.00233.x

[58] Randy Stein and Alexander B. Swan. 2019. Evaluating the validity of Myers-Briggs Type Indicator theory: A teaching tool and window into intuitive psychology. Social and Personality Psychology Compass 13, 2, e12434. DOI:https://doi.org/10.1111/spc3.12434

[59] Sarah Theres Völkel, Ramona Schödel, Daniel Buschek, Clemens Stachl, Verena Winterhalter, Markus Bühner, and Heinrich Hussmann. 2020. Developing a Personality Model for Speech-based Conversational Agents Using the Psycholexical Approach. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–14. DOI:https://doi-org.ezproxy.pratt.edu/10.1145/3313831.3376210

[60] Ilaria Torre, Jeremy Goslin, and Laurence White. 2019. If your device could smile: People trust happy-sounding artificial agents more. Computers in Human Behavior 105, 106215. DOI:https://doi.org/10.1016/j.chb.2019.106215

[61] Peter Wright. 1974. The harassed decision maker: Time pressures, distractions, and the use of evidence. Journal of Applied Psychology 59, 5 (1974), 555–561. DOI:https://doi.org/10.1037/h0037186

[62] Dan Wu, Shaobo Liang, Jing Dong, and Jin Qiu. 2018. Impact of Task Types on Collaborative Information Seeking Behavior. Libri 68, 3, 231–245. DOI:https://doi.org/10.1515/libri-2017-0051

[63] Mao Xuetao, François Bouchet, and Jean-Paul Sansonnet. 2009. Impact of agent's answers variability on its believability and human-likeness and consequent chatbot improvements. In Proceedings of the Symposium Killer Robots vs. Friendly Fridges—The Social Understanding of Artificial Intelligence (AISB 2009), 31–36.

[64] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting Virtual Agents: The Effect of Personality. ACM Trans. Interact. Intell. Syst. 9, 2–3, Article 10, 36 pages. DOI:https://doi.org/10.1145/3232077