Make your own

The potential of chatbot customization for the development of user trust

Wald, R.; Heijselaar, E.; Bosse, T.

DOI
10.1145/3450614.3463600

Publication date
2021

Document Version
Final published version

Published in
UMAP '21

License
Article 25fa Dutch Copyright Act

Make your own

The Potential of Chatbot Customization for the Development of User Trust

Rebecca Wald
Amsterdam School of Communication Research, University of Amsterdam
r.wald@uva.nl

Evelien Heyselaar
Behavioural Science Institute, Radboud University
e.heijselaar@maw.ru.nl

Tibor Bosse
Behavioural Science Institute, Radboud University
t.bosse@ru.nl

ABSTRACT

Interacting with chatbots has become ubiquitous nowadays. Nevertheless, conversational agents often remain unable to reliably succeed in social contexts, which negatively influences users' experience and prevents them from exploiting the technology's full potential. To improve user experience and subsequent trust-formation in chatbots only very little attention has been paid to the active involvement of the user and with that to customization options. Employing a preregistered experimental 1x2 between-subjects study design (N = 171) this study explores an alternative approach to the typical one-chatbot-fits-all solution and investigates the potential of active user-based chatbot customization for the development of trust in chatbots. While customization had no direct effect on trust, anthropomorphism was identified as a significant mediator. The chatbot's interpersonal communicational competence was not affected by customization, yet it did predict trust. Exploratory analyses of participants' feedback point towards the importance of individual differences between users and generally show a positive impact of customization on the overall chatbot experience.

CCS CONCEPTS

• **Empirical studies in HCI**; • **Interaction design theory, concepts and paradigms**;

KEYWORDS

Anthropomorphism, Chatbots, Interpersonal Communicational Competence, Trust, User Experience Design

ACM Reference Format:
Rebecca Wald, Evelien Heyselaar, and Tibor Bosse. 2021. Make your own: The Potential of Chatbot Customization for the Development of User Trust. In *Adjunct Proceedings of the 29th ACM Conference on User Modeling, Adaptation and Personalization (UMAP '21 Adjunct), June 21–25, 2021, Utrecht, Netherlands.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3450614.3463600

1 INTRODUCTION

Promised to be of great economic and societal value [19], text-based conversational agents, or chatbots, have become omnipresent in users' "personal digital universe" [16, p. 8]. Yet, despite their rising presence, with the expectation to communicate with the chatbot as if it were another human, users often seem to overrate the performance of conversational agents [19; 25]. In consequence, users' frustration rises [19] and their likelihood for future chatbot encounters decreases [25; 18]. This becomes cumbersome as, eventually, a gap emerges between the chatbot and its user prohibiting subsequent trust-formation with chatbots and, thus, further exploitation of the technology's potential.

Up until now, most strategies that are aimed at improving users' experience with chatbots, thereby bridging the gap between both entities, originate from the developers' side. Less attention has been paid to users' active involvement and co-creation procedures (i.e., through customization), although a handful existing findings already provide first indicators for positive effects [12]. On this basis, the present study attempts to strengthen the scientific understanding of trust in Human-Computer Interaction (HCI) by exploring user-based chatbot customization as an alternative approach to the typical one-chatbot-fits-all solution. The research question guiding this investigation is as follows: *To what extent does user-based chatbot customization influence the user's trust in the chatbot?*

1.1 Theoretical background

1.1.1 Trust components. In the Human-Human Interaction (HHI) context trust generally functions to build and maintain relationships [35]. Under the assumption of the Computers Are Social Actors paradigm [23], which understands humans as responding to computers in a social manner by applying human-specific concepts to human-appearing but, eventually, non-human entities, it can be assumed that for the HCI context this interpersonal construct of trust persists in a similar manner. Respective literature that further tests this assumption, however, remains scarce. Among those limited research findings two essential components seem to tap more closely into the construct of trust between a human and computational entity: the technology's *competence* and *anthropomorphic* qualities.

Since chatbots are increasingly designed to fulfill interpersonal communicational tasks, for instance in the form of engaging in casual small-talk, their *communicational* competence gains great importance. In light of this, Skjuve and Brandzaeg [29] have adapted the Interpersonal Communicational Competence (ICC) scale to the evaluation of interpersonal chatbot encounters, with ICC referring to the perceived ability of a chatbot "to manage interpersonal relationships in communication settings" (Rubin and Martin, 1994, p. 33). First findings show that ICC impacts users' general willingness to engage with social chatbots [6] and ICC is shown to be positively associated with a chatbot's greater social presence [29].

UMAP '21 Adjunct, June 21–25, 2021, Utrecht, Netherlands

ACM ISBN 978-1-4503-8367-7/21/06…$15.00
https://doi.org/10.1145/3450614.3463600

Social presence, in turn, is known to improve the user's communicative experience and the better this experience, the more likely the user is to perceive the chatbot as credible and trustworthy [8].

Anthropomorphism generally describes the attribution process of visual as well as cognitive human characteristics to non-human entities [33]. Empirical evidence exists for a positive relationship between anthropomorphism and trust in virtual non-human entities when measured through self-reports [18]. Particularly, Cowell and Stanney [9] found chatbots' visual characteristics of gender, ethnicity, and age to be especially influential for the degree of perceived anthropomorphism, while research in the related field of intelligent automobiles further concludes that specific features about the technology's identity can be crucial to strengthen feelings of trust (33Additionally, Waytz et al. [33]'s findings suggest that aspects of anthropomorphism can moderate users' perception of technological competence, suggesting that an interaction between competence and anthropomorphism is very likely.

1.1.2 Customization effects. Arguably, the perceived competence and anthropomorphic qualities of chatbots depend on their respective *functionality* and *identity*. To date, a chatbot's functionality and identity are usually created by a developer, hence, they remain the same for every user [11]. Existing research, however, emphasizes the important role of users' individual differences that significantly affect perceptions of chatbots (Nass and Lee, 2000). A one-chatbot-fits-all solution might therefore not be optimal. In the field of gaming research, the implementation of an alternative approach, namely in form of game avatar customization, is already a common feature. In fact, first indicators exist showing positive effects of such customization on user experience [12]. From a behavioral science perspective, the theoretical argument for this effect is two-fold: One, the active involvement of the user is in line with the Uses and Gratifications (U&G) theory [20], which suggests that users "actively" choose a medium to fulfil their individual needs [6]. Hence, every user can have an individual use intention. Two, the impact of co-creation can be explained by the so-called IKEA effect [21], which suggests that customers show greater willingness to consume, engage with, and benefit from products they self-create. Previous work in HCI concludes that an increase in the perceived identity disclosure of a virtual character can have a beneficial impact on the feelings of control and trust on the user side [30]. The IKEA effect could therefore form a fruitful explanation for an increase in user experience and potential trust-formation when users are able to self-create their chatbot.

To our knowledge, one other study has previously explored such customization effects to enhance user trust in chatbot encounters. Xiao et al. [34] utilized a 2×2 balanced, between-subjects design, with a customizable or generic and a well-qualified (i.e., professional appearance) or poorly qualified (i.e., cartoon-like appearance) conversational agent. Participants who underwent the customization eventually rated the agent "as more lik[e]able, more trustworthy and more useful" (p. 1299) and this was generally more influential than the agent's degree of qualification. Yet, since participants were told that reporting on their own personal preferences would allow the system to later determine the best suited chatbot 'for' them, only the illusion of customization was created and tested. Under the assumption that if an illusion can show significant effects on how users experience customized chatbots, this study aims to replicate this finding while implementing an *active* user-based customization procedure. In addition, we strive to advance the theoretical argumentation and to update this line of research under the premise of the rapid development in technology and society nowadays. To do so, we focus on the customization of a disembodied chatbot in the context of an unspecific one-time encounter with customization being defined as "the degree to which [the] technology [. . .] or service can be created, selected, or changed to comply with user preferences" [31].

1.2 Conceptual model and hypotheses

Based on the illustrated theoretical framework and existing literature, we propose the following conceptual model (Figure 1) and confirmatory hypotheses (see Table 1 for exact wording).

2 METHODS

2.1 Design

We used a preregistered (10.17605/OSF.IO/4MDTR) experimental 1x2 between-subjects study design, which was officially approved by the Radboud University's ethics committee (ECSW-2019-151). Respective data management guidelines were followed. The study utilized an online survey administered via Qualtrics, and a web-based chatroom created through an online website building platform. Initially, the experiment was designed to be conducted at the research laboratory on the Dutch campus, however, due to the unforeseen COVID-19 outbreak at the time of data collection the study was restructured into an online experiment half-way through. Comparison of the two data-collection strategies revealed no significant differences in the study's outcome variables.

2.2 Procedure

After giving consent to participate in this study, demographic and personal information (e.g., personal interests) of participants were collected in the first part of the online questionnaire. Depending on the assigned condition, the questionnaire further asked the participant to customize their chatbot along multiple-choice options (i.e., male/female, brighter/darker skin tone, blue/green/brown eyes, blond/red/dark hair, English/Dutch/German nationality, personal interest for movies/music/books/sports/food, and an open field for the chatbot's name) or it introduced them to the generic chatbot. A three-minute initialization process was implemented for the set-up of the respective online-chatroom, enabling participation of only one participant at a time. After being directed to the chatroom via an external URL each participant had a ten-minute conversation with the chatbot about personal interests, which the pre-test identified as a reasonable duration and conversation topic. As no suitably advanced and customizable chatbot existed upon data collection, we used the openly available chatbot Mitsuku [24], whose software imitates human conversation in a non-specific context [1] based on information retrieval of free user inputs. To properly integrate the chosen features in the customization condition, we used a Wizard-of-Oz approach [10] through which the researcher – as the wizard – altered Mitsuku's messages to fit the participant's customization choices. These changes were made along pre-defined standardized scripts (see material on OSF). The response time in

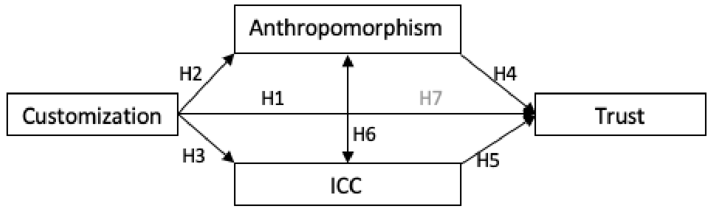

Figure 1: Conceptual Model. Note: Hypothesis H7 illustrates the mediation hypothesis of customization on Trust via Anthropomorphism and ICC.

the control condition was kept as similar as possible to one in the customization condition. After ten minutes, the chatbot thanked the participant for the conversation and gave instructions on how to return to the online-questionnaire, which assessed the participant's chatbot experience via self-reports consisting of responses on the measurement scales.

Finally, to ensure whether the chatbot interaction was believable and, thus, whether the answers on the self-report scales were valid, a manipulation check was implemented in form of a funnel debrief based on open questions (i.e., In this experiment there were two groups. One was talking to a chatbot and the other one was talking to a human. Which one do you think you belonged to? If choosing human group: To improve the experiment please indicate at what point you realized that you were talking to a human? If choosing chatbot group: Why did you believe that your conversation partner was a chatbot?). Answers were coded by two independent coders revealing a sufficient inter-rater reliability Kappa score of .96 [32]. In total, eleven participants failed the manipulation check.

2.3 Participants

Prior to data collection a G*Power-Analysis [13] was conducted. Based on related work by Kulm and Kopp (Kulm and Kopp, 2019), who detected an effect size of $\eta_p{}^2$ = .05 for self-reported trust in virtual agents, we calculated a total sample size of N = 249 with an alpha of .05 and a power of 95% for our anticipated statistical tests, i.e., ANOVA with fixed effects, main effects, and interactions for a between-subject design. The final sample, which was recruited through the university's participant pool as well as through social media platforms using a convenience sampling strategy, consisted of 171 valid responses of participants who were non-dyslexic, without any mental health diagnosis and able to understand and write in English. Even though this final sample size lay below the previously computed sample size, it still exceeded the 85% power level and was found to be adequate. Among those 171 participants, 123 were female (72%); one participant declared to neither belong to the female nor the male category. Split by the two groups, 81 participants were assigned to the experimental condition (74% females; 51% collected in the lab) and 90 participants to the control condition (70% females; 41% collected in the lab). In sum, 79 (46%) of the responses were generated in the laboratory. All participants were reimbursed by credit points or five Euros for their approximately 25 minutes participation.

2.4 Measures

The chatbot's competence was measured via the ICC scale by Skjuve and Brandzaeg [29]. Since Mitsuku was, up until data collection, not able to effectively communicate through emojis, items that directly related to such communicative aspects were excluded. Furthermore, items that referred to a dominant chatbot role to lead, negotiate or evaluate the conversation topics – typical qualities for custom support chatbots – were excluded, because the participant was asked to take the active lead and because Mitsuku's software is programmed to be open and non-directed [17]. Each item was measured on a five-point Likert scale, ranging from 1 = *Not true at all* to 5 = *Very true.* The chatbot's degree of anthropomorphism was measured through the anthropomorphism scale of the Godspeed Questionnaire [3] along a five-point differential. The fifth item, referring to movements of the chatbot, was judged to be unsuitable for our investigation of a disembodied chatbot. To measure the participant's trust in the chatbot we used the trust scale by Bickmore et al. [5] measured on a seven-point Likert scale ranging from 1 = *Not at all* to 7 = *Completely.* Finally, in correspondence with existing work on user affinity with conversational agents [25] we included a measurement of the participants' frequency of previous chatbot interaction (i.e., daily, about once a week, about once a month, usually never, today was the first time).

2.5 Analyses

For the statistical analyses we used the software R for statistical computing [26]. The preregistered linear model, controlling for participants' frequency of chatbot interaction, was performed using the lme4 R-package version 1.1-21 [2] with the default setting of treatment contrasts using the preregistered linear model structure. Next to the confirmatory analyses, the participants' open feedback responses were examined in an exploratory manner to formulate valuable suggestions for future chatbot development.

Factor scores were computed for each participant's answer on the anthropomorphism and ICC scale using the R-packages corpcor version 1.6.9 [28], GPArotation version 2014.11-1 [4], and psych version 1.8.12 [27]. A series of respective assumption checks (i.e., Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy, Bartlett's test of Sphericity, and communalities) were performed on the data. The first Principal Component Analysis (PCA) using orthogonal rotation (e.g., varimax) revealed that for the anthropomorphism scale one component had an eigenvalue over Kaiser's criterion of 1 and explained 62% of the variance. The reliability of the anthropomorphism scale revealed good reliability (α = .79). The second PCA using orthogonal rotation (i.e., varimax) was conducted for the items of the ICC scale. Due to the relatively small sample size and low correlation coefficients of items 5.1 and 5.2 (on average $< .10$), there was reason not to trust the adequate KMO measure of .82 and the significant results of the Bartlett's test of Sphericity (X^2 (136) = 872.36, $p < .001$). Removing those two items showed an improvement of the correlation matrix and respective coefficients (KMO = .83; X^2 (105) = 834.91, $p < .001$). Four components had eigenvalues over the Kaiser's criterion of 1 and in combination explained 60% of the variance. Scale reliability was good for components performance (α = .75), emotional transparency (α = .71), and social relaxation (α = .81). Cronbach's alpha for Closeness (α = .23) revealed poor reliability.

3 RESULTS

3.1 Confirmatory analyses

The experimental group reported higher scores for trust (M = 3.68, SD = 1.59) than the control group (M = 3.32, SD = 1.53). However, a Mann-Whitney U test, which was appropriate to analyze the variable's non-normal distribution, showed this difference was not statistically significant, U = 3204.00, p = .163, r = 0.11, which provided no support for hypothesis H1. In contrast, results provided support for the impact of customization on anthropomorphism (H2), as an independent samples t-test showed a significant difference between the experimental (M = 0.21, SD = 1.03) and control group (M = -0.19, SD = 0.93), t(162.09) = 2.68, p = .008, 95% CI [0.16, 0.66], Cohen's d_s = 0.42. The ICC-scores for *Emotional Transparency*, *Closeness*, and *Social Relaxation* were higher in the experimental group than in the control group. Independent samples t-tests, however, revealed no statistical significance of these differences. Hypothesis H3 was therefore rejected.

The regression model explained 42% of the variance; a collective significant effect was found; R^2 = 0.42, F(16, 154) = 8.564, $p < .001$. The model results provide support for hypotheses H4 and H5, as anthropomorphism (β = 0.373, t = 2.798, p = .005) as well as all ICC components (*Performance*: β = 0.287, t = 2.475, p = .014; *Emotional Transparency*: β = 0.458, t = 4.220, $p < .001$; *Closeness*: β = 0.340, t = 3.401, $p < .001$; *Social Relaxation*: β = 0.475, t = 4.647, $p < .001$) could be identified as significant predictors of trust. The model results further revealed a significant interaction between anthropomorphism and the ICC component *Closeness* (β = -0.218, t = -2.305, p = .022). The interaction hypothesis (H6) was therefore partially supported.

Following the new school of mediation analysis [15] that does not necessarily require a significant direct effect between the independent and dependent variable (here customization on trust), a significant mediation effect of anthropomorphism was found. The regression coefficient between customization and anthropomorphism (β = -0.212, t = -2.812, p = .005) was significant and results confirmed that anthropomorphism had an effect on trust when controlling for customization. Unstandardized indirect effects were computed for each of 1000 bootstrapped samples, and the 95% confidence interval was computed by determining the indirect effects at the 2.5th and 97.5th percentiles. The bootstrapped unstandardized indirect effect was .344, 95% CI [0.10; 0.58]. No such mediation could be found for any of the ICC components. Hypotheses H7 was therefore partially supported. A summary of the study findings is shown in Table 1

3.2 Exploratory analyses

When analyzing the open feedback responses generated through the funnel-debrief questions using a thoroughly developed codebook (see materials on OSF), participants most frequently reported arguments concerning the chatbot's poor conversational quality, followed by criticism of the chatbot's understanding and use of language. Fewest critical references were made to the chatbot's identity. Interestingly, fundamental differences between participants' evaluation of the chatbot were found in the form of highly positive as well as negative evaluations of the same aspect (e.g., reference to conversational history, one-sided conversational flow). With regards to significant differences in those feedback responses a Pearson's chi-square test of independence showed that the association between condition and positive feedback was statistically significant, X^2 (1, N = 171) = 10.74, p = .001, Cramer's V = 0.25, with participants in the customization condition expressing more positive feedback than the control group.

4 DISCUSSION

The aim of this study was to explore an alternative approach to the typical one-chatbot-fits-all solution by investigating the potential of active user-based chatbot customization for the development of trust in chatbots. Trust in chatbots appears to be dependent on the anthropomorphic impression and communicational competence of the chatbot. Considering the significant negative interaction between anthropomorphism and the ICC component of closeness we have ground to think that those two measures capture parts of the same trust-predicting construct, which leads to a reduction of their individual effects. In sum, the finding of Xiao et al. [34] that a customized chatbot is perceived as more trustworthy, was partially replicated, however, only under the condition that the chatbot was perceived as more anthropomorphic.

Theoretically speaking, since the anthropomorphic cues used in the customization condition were the same in the control condition, the mechanism of the IKEA effect [21] seems to indeed be fruitful when aiming to improve users' valuation of chatbots. This extends the view of [34] in a meaningful way: In terms of user trust human-like might only be one side of the coin; self-made could be the other one. This finding can be particularly useful for real-world chatbot encounters that aim to create a human-like conversation setting and a more intimate relationship between the chatbot and its user (e.g., health/therapeutic contexts).

The participants' open feedback responses further showed that, as mentioned in the literature, chatbots do not yet fully meet user expectations in terms of their ability to use human language and grasp conversational context. Here, many individual differences were found. While some participants were impressed by the chatbot's performance, for others it created an eerie sensation, which the scientific literature calls the uncanny valley of mind effect [22]. It is triggered if a non-human agent appears and acts too human-like, which has mostly been found for chatbots with animated avatars [7]. The threshold for this effect seems to be individually different, which is why future chatbot development should aim to leave users

Table 1: Summary of findings

	Hypothesis	Finding
H1	Users who customize the chatbot will report greater trust in the chatbot after interacting with it than users who interact with a generic chatbot.	Rejected
H2	Users who customize the chatbot will report higher degrees of anthropomorphism after interacting with it than users who interact with a generic chatbot.	Supported
H3	Users who customize the chatbot will report higher ICC scores after interacting with it than users who interact with a generic chatbot.	Rejected
H4	The ascribed degree of anthropomorphism of the chatbot is predictive of user trust in the chatbot.	Supported
H5	The perceived ICC of the chatbot is predictive of user trust in the chatbot.	Supported
H6	The ascribed degree of anthropomorphism and the perceived ICC interact with each other.	Partially supported
H7	The relationship between customization and trust is mediated by the degree of ascribed anthropomorphism and perceived ICC.	Partially supported

more room to choose how human-like they want their chatbot to be.

Although this research made its mark by investigating the effects of the novel feature of chatbot customization on user trust, it also had its limitations. First of all, the chosen scenario for the chatbot encounter and the methodological approach are likely to have resulted in many degrees of freedom between each individual chatbot encounter. To counteract this shortcoming, we advise future scholars to design an application that integrates both, customization and chatroom, in one automated interface. Secondly, given the small sample size of this study, we see a necessity for more thorough scientific engagement with the adjusted ICC scale by Skjuve and Brandzaeg [29] to properly validate the measurement of communicational competence for different contexts of chatbot encounters. Thirdly, the limited scope of this work did not allow closer examination and storing of the chatbot conversations. Since self-reports only capture explicit attitudes of participants that are likely to be distorted by reflective cognitive processes [16], especially when it comes to a complex mental construct such as trust-formation, measures that focus on behavioral data (e.g., linguistics) could shed further light onto the effects of customization.

5 CONCLUSION

Since a growing number of tasks in everyday life are executed with the help of conversational interfaces today, building and maintaining adequate trust in chatbots has become highly important. As a valuable contribution to the growing body of HCI research, this study showed that the alternative approach of active user-based chatbot customization has the potential to affect trust-formation in chatbots via anthropomorphic cues. Future chatbot development should concentrate more on user-driven design features that allow the user to have more control over who they interact with. Human-like might only be one side of the coin; self-made could be the other one in order to put AI-technology at the service of society going forward.

REFERENCES

[1] Abdul-Kader S. A. & Woods, J.(2015). Review on chatbot design techniques in speech conversation systems. International Journal of Advanced Computer Science and Applications, 6(7), 72–80. https://doi.org/10.17148/iarjset.2018.596

[2] Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. https://doi.org/10.18637/jss.v067.i01

[3] Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics, 1(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[4] Bernaards, C. A., & Jennrich, R. I. (2005). Gradient Projection Algorithms and Software for Arbitrary Rotation Criteria in Factor Analysis. Educational and Psychological Measurement, 65, 676-696.

[5] Bickmore, T., Vardoulakis, L., Jack, B., Paasche-Orlow, M. (2013). Automated Promotion of Technology Acceptance by Clinicians Using Relational Agents. In Aylett R., Krenn B., Pelachaud C., Shimodaira H. (eds) Intelligent Virtual Agents. IVA 2013, 68-78.

[6] Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. In Kompatsiaris *et al.* (Eds.). Internet Science: 4th International Conference, INSCI 2017, Thessaloniki, 377–392. https://doi.org/10.1007/978-3-319-70284-1_30

[7] Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2019). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. Future Generation Computer Systems, 92, 539-548.

[8] Chung, M., Ko, E., Joung, H., & Kim, S. J. (2018). Chatbot e-service and customer satisfaction regarding luxury brands. Journal of Business Research, 1–9. https://doi.org/10.1016/j.jbusres.2018.10.004

[9] Cowell, A. J., & Stanney, K. M. (2003). Embodiment and Interaction Guidelines for Designing Credible, Trustworthy Embodied Conversational Agents. In Intelligent Virtual Agents. 4th International Workshop, IVA 2003 (pp. 301–309). https://doi.org/10.1016/s0097-8493(03)00043-8

[10] Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies - why and how. Knowledge-Based Systems, 6(4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-N

[11] Daniel, F., Zaccaria, V., Matera, M., & Dell'orto, A. (2018). Toward truly personal chatbots. on the development of custom conversational assistants. ACM/IEEE. https://doi.org/10.1145/3195555.3195563

[12] Dolgov, I., Graves, W. J., Nearents, M. R., Schwark, J. D., & Brooks Volkman, C. (2014). Effects of cooperative gaming and avatar customization on subsequent spontaneous helping behavior. Computers in Human Behavior, 33, 49–55. https://doi.org/10.1016/j.chb.2013.12.028

[13] Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power anal- ysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

[14] Følstad, A., Skjuve, M., & Brandtzaeg, P. B. (2018). Different chatbots for different purposes: Towards a typology of chatbots to understand interaction design. Proceedings of the international workshop on chatbot research CONVERSATIONS 2018, St. Petersburg, Russia.

[15] Hayes, A. F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach (1st ed.). New York: The Guilford Press.

[16] Heyselaar, E., & Bosse, T. (2020). Using theory of mind to assess users' sense of agency in social chatbots. In: A. Følstad *et al.* (Eds.): CONVERSATIONS 2019, LNCS 11970, pp. 79–93, 2020. https://doi.org/10.1007/978-3-030-39540-7_6

[17] Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth and uHealth, 6(11), e12106.

[18] Kulms, P., & Kopp, S. (2019). More human-likeness, more trust? The effect of anthropomorphism on self-reported and behavioral trust in continued and interdependent human-agent cooperation. MuC '19, September 8–11, 2019, Hamburg, Germany, 31–42. https://doi.org/10.1145/3340764.3340793

[19] Luger, E., & Sellen, A. (2016). "Like having a really bad pa": The gulf between user expectation and experience of conversational agents. CHI'16, May 07 - 12, 2016, San Jose, CA, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288
[20] McQuail, D. (1994). The rise of media of mass communication. In D. McQuail (Ed.), Mass communication theory: An introduction (pp. 1–29). London: Sage.
[21] Mochon, D., Norton, M. I., & Ariely, D. (2012). Bolstering and restoring feelings of competence via the IKEA effect. International Journal of Research in Marketing, 29 (2012), 363-369.
[22] Mori, M. (1970). The Uncanny Valley. Energy, 7(4), 33–35.
[23] Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. Conference on Human Factors in Computing Systems - Proceedings, 72–78.
[24] Pandorabots (2019). Meet Mitsuku. URL: https://www.pandorabots.com/mitsuku/
[25] Piccolo, L. S. G., Mensio, M., & Alani, H. (2018). Chasing the Chatbots? Directions for Interaction and Design Research. Proceedings of the international workshop on chatbot research CONVERSATIONS 2018, St. Petersburg, Russia.
[26] R, Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
[27] Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.
[28] Schafer, J., Opgen-Rhein, R., Zuber, V., Ahdesmaki, M. A., Duarte Silva, P., & Strimmer, K. (2017). corpcor: Efficient Estimation of Covariance and (Partial) Correlation. R package version 1.6.9. https://CRAN.R-project.org/package=corpcor
[29] Skjuve, M., & Brandzaeg, P. B. (2018). Measuring user experience in chatbots: An approach to interpersonal communication competence. Proceedings of the international workshop on chatbot research CONVERSATIONS 2018, St. Petersburg, Russia. https://doi.org/10.1007/978-3-030-17705-8_10
[30] Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of Human-AI Interaction (HAII). Journal of Computer-Mediated Communication, 74-88. https://doi.org/10/ggvvq
[31] Teng, C. (2010). Customization, immersion satisfaction, and online gamer loyalty. Computers in Human Behavior, 26(6), 1547–1554. https://doi.org/10.1016/j.chb.2010.05.029
[32] Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistics. Family Medicine, 37(5): 360-3.
[33] Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. Journal of Experimental Social Psychology, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005
[34] Xiao, J., Stasko, J., & Catrambone, R. (2007). The role of choice and customization on users' interaction with embodied conversational agents: Effects on perception and performance. Proceedings of the conference on Human Factors in Computing Systems, 1293–1302. https://doi.org/10.1145/1240624.1240820
[35] Yen, C., & Chiang, M. C. (2020). Trust me, if you can: A study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image evidence and self-reported assessments. Behaviour & Information Technology, https://doi.org/10.1080/0144929X.2020.1743362