Design for the User You Want, Not the User You Have?

Anna-Maria Meck
Ludwig-Maximilians-Universität München
Munich, Germany
a.meck@campus.lmu.de

Justin Edwards
Diego Garaialde
Marion Bartl
Philip Doyle
University College Dublin
Dublin, Ireland
justin.edwards@ucdconnect.ie
diego.garaialde@ucdconnect.ie
marion.bartl@ucdconnect.ie
philip.doyle1@ucdconnect.ie>

Leigh Clark
Swansea University
Swansea, Wales
l.m.h.clark@swansea.ac.uk

Abstract
Conversational User Interfaces (CUIs) have seen little diversity in the types of voices that are represented in their design. This has led to speculation that norms for synthetic voices will create and reinforce stereotypes, leading to biases against both those whose voices are represented by CUIs and those whose voices are not. We propose a solution for this problem - CUI designers should force users to interact with CUIs which have a more diverse selection of voices, and CUI researchers should focus on such CUIs rather than those that reinforce norms. Here, we lay out opportunities and challenges in addressing social biases and stereotypes through the way we design and research CUIs.

Author Keywords
conversational user interfaces, voice user interfaces, bias, gender bias, stereotypes

Defining the problem
Now that CUIs are becoming relatively competent from a technical perspective, designers and researchers alike are increasingly engaging in topics around user experience, building ‘natural’ conversations, and considerations around ethics and ethical design. These considerations are especially important as the number of CUI users reaches billions, making CUIs an ubiquitous technology [13]. In 2000, Nass and Moon coined the computers are social actors

(CASA) paradigm, stating that humans treat computers, not as humans, but as some alternative type of social actor, "applying gender stereotypes to computers and ethnically identifying with computer agents" [21]. Furthermore, users employ learned social behaviours from human-human interactions like politeness or reciprocity to human-computer interactions[21]. Individuals hence anthropomorphize computers, robots, and CUIs and treat them as social entities. Research shows that this anthropomorphization facilitates adoption and usage of new technologies [35, 36] which is why designers rely heavily on making CUIs even more human-like by giving them a personality, a voice, human-like appearances and a gender. Gender can be expressed through the given name (e.g. "Siri" or "Alexa"), the fundamental frequency of a voice, an avatar (e.g. Microsoft's Cortana), pronouns, and even speaking-style [32, 8] . While big brands claim to not assign gender to their CUIs [1], a study by Feine et al. found that the majority of chatbots do display gender-specific cues, in 77.56% of cases female ones [8].

The problem of how harmful gender stereotypes impact society is an ongoing issue that still needs to be overcome. Gendered CUIs open up a new front line for this problem, with research showing harmful biases and gender stereotypes are also found in users' language towards CUIs [34, 2]. With a current design paradigm that deems the user as "king/queen", CUIs have been expressly designed to serve their masters' every need. If an error occurs, a CUI will not blame the user but instead take responsibility for the issue and sometimes apologise for messing up. If a CUI is being talked to impolitely or even insulted, it will most likely not stand up for itself but respond nonchalantly and evasively (in 2017, Alexa's response to "You're a bitch" was "Well, thanks for the feedback.", see [9]). Even if a person talks to a CUI in an impersonally demanding, aggressive or hostile manner, it will still tend to a user's command with no agency of it's own. This creates a strong power dynamic between users and mostly female CUIs [12] that are obliging, subservient, and okay with being treated poorly. Furthermore, they are available via a simple press of a button, a quick "Hey. . . " (e.g. Siri) or just a name (Alexa). The fear is that these patterns may translate back to the way we engage with women in human-human communication: "the more that culture teaches people to equate women with assistants, the more real women will be seen as assistants – and penalised for not being assistant-like." (p.106 of [34] citing [14].) Researchers should deeply consider the ethics of gendered voices when designing conversational interfaces. Through a combination of the Q and A dynamics of dialogue, the asymmetrical relationship between system and user, and the voices represented through these systems, present significant potential for reinforcing underlying negative societal biases. This may mean changing current design paradigms and making CUIs more diverse, not only in terms of gender, but also across other identity dimensions like regional accents, markers of social class, and age. How will CUI users react to a shift like this? Challenging CUI design stereotypes would ideally challenge a users' preconceived notions of what CUIs should sound like and say, a process that may create user discomfort and induce dissatisfaction with the product. This workshop paper aims to explore the opportunities and challenges that would result from forcing users to challenge their stereotypes around CUIs.

Opportunities of forcing CUI users to challenge stereotypes

Ongoing research and discussion of CUIs [34, 3, 32] has speculated that harmful stereotypical behaviour in human-computer interaction may encourage similar behaviour in human-human interaction. We propose a constructivist addendum to this notion: that challenging harmful stereo-

typical behaviour in human-computer interaction can also translate back to human-human interaction. We suggest that by exposing users' to a more diverse set of CUI voices, we might open avenues for this diversity to become part of their day-to-day communication patterns. Designed in this way, CUIs could bring users in touch with identities they may not experience in their everyday lives and thus create openness towards humans that hold these identities. The idea is fueled by a study conducted by Lopatovska et al., where the authors show that while users preferred female and male voices over non-binary voices, participants also "expressed their ideal agent as being non-binary". This led the authors to speculate about "a cultural shift in user expectations (at least in theory)" [17], coming to the conclusion that this theoretically perceived cultural shift is well worth investing in. And it is, insofar as we rarely see, or rather hear much variation in synthesis from norms established by the likes of Apple, Google, and Amazon, whilst benefits and user preferences really haven't been greatly explored. In research that does examine this issue, Scott et al., show that although people expressed initial reservations about hearing weather reports read using regionally accented text-to-speech (TTS) voices, upon actually hearing them, they felt emotional connection to the accented voice [28]. Yet, work of this nature is very limited. Hence, it may be the case that users are a lot more open to a diverse set of CUIs than we may know at this point - simply because of a lack of research in this area. It can even be argued that giving users a choice when it comes to the identity of their CUI will equip them with an elevated sense of agency. Shneiderman's 7th Golden Rule of Interface Design states that designers need to support internal locus of control as users "strongly desire the sense that they are in charge of the system (...)." Systems need to be designed in a way that makes "users the initiators of actions rather than the responders" [30] Giving users the opportunity to choose between different CUI identities may achieve just that.

The lack of research on CUIs with more diverse social identities, in terms of more varied regional accents, expressions of gender, and linguistic markers of social class results in established norms for voices of CUIs rarely being challenged. These norms, which likely stem from earlier voices used for IVR telephony customer service agents, have also been strengthened by the commercial popularity of conversational interfaces like Amazon's Alexa, Apple's Siri, Google Assistant, and Microsoft's Cortana. These limited identities have also been inadvertently reinforced by academic research on CUIs as academics seek to better understand both the CUIs that people already use (in works like [38, 37, 24, 5]) as well as understanding imagined or prototypical CUIs that share characteristics with existing CUIs (in works like [7, 11, 29]). By focusing research on this limited range of CUIs, academia fortifies ideas about what voices are 'suitable', furthering the negative stereotypes and deeper entrenching biased associations that emerge from people's interactions with CUIs, whose voices only sound a particular way. In a sense, this issue resonates with and contributes to "othering" of certain groups due to the narrow datasets language models are built on, which leads to poorer ASR efficacy for women, children and other minority groups [10, 26]. In a similar fashion to arguments made about the classic BBC voice [16], rather than being merely a representational problem, using a normative voice for CUIs divides people into social in- and out-groups containing, "people with voices appropriate for a CUI" and the "other". The reinforcement of this sort of social othering through academic research, termed epistemological violence [33] is said to contribute to broader social mistreatment of outgroups and entrench the status of "other" beyond the bounds of social science research. Following Sug-

arman, the tendency of academia to endorse and reinforce neoliberal ethics - in this case, allowing private companies and the free market of their customers to decide what a CUI can sound like - fortifies those very ethics in society, indirectly endorsing the ethical principles and sociopolitical order under which research is conducted [31].

Challenges of forcing users to challenge stereotypes

Still, issues with forcing users to challenge stereotypes need mentioning. The gulf of evaluation is a well-established phenomenon in HCI [22, 18] and suggests violations of expectation are a key cause of limited use and abandonment. As most CUIs today are female representations of assistant personas, users know what to expect when using them. Adoption of CUI voices that deviate from this codified norm could potentially be arduous for users. This notion is supported in work from Pawlik who found that "gendered anthropomorphic chatbot design cues yield significant positive impact of the performance expectancy towards the behavioural intention to use chatbots" [23]. A critical risk for CUI designers and practitioners is that CUIs with voices that their users find inappropriate might lead to greater abandonment of CUIs altogether. This is particularly a concern given research indicating that judgements people make about CUI capabilities are strongly influenced by voices based on their voices, and that these judgements subsequently impact interaction behaviour [20, 4]). As such, attempting to reduce the potential for causing social harm through CUI design may create significant usability issues, for systems already struggling to live up to their hype. Further, some evidence from Human Robot Interaction (HRI) implies that fears about negative consequences resulting from gender biases in HCI with nonhuman agents may be overblown. The work suggests many established gender stereotypes in human-human interaction fail to emerge in the context of HRI studies [25]. For this reason, it is important that research on this topic is incremental and that risks and benefits are considered carefully, with recognition for the fact that human-human interactions and interactions with CUIs are not necessarily analogous.

While we encourage CUI users to challenge the stereotypes, CUI researchers and designers face the challenge of deciding exactly how to do this. If, on the one hand, we follow established HCI guidelines of giving users choices by including a broadly diverse selection of voices for their CUIs, there is a distinct risk that users will ignore the options and fall back on the norms that have already been established. In order to mitigate that risk, designers may decide to violate this principle of choice, instead proscribing particular voices for their CUIs. This becomes problematic however, so long as CUIs remain in the subservient role of the assistant or the butler; responsive things with little ability for proactivity [6]. In this way, designers may end up reinforcing negative representations that these subservient roles ought to be carried out by an agent with a voice with characteristics in common with people who lack social power. For this reason, forcing voices on users might best be saved for a future, when the functionality of these systems has matured and the roles they might be capable of fulfilling are more clear. A time when we can "move beyond the butler" [19] and adopt different interaction metaphors for the social roles of CUIs, such as teammates, collaborators, or colleagues. In fact, it may be possible to completely subvert this interaction metaphor and sidestep problems of promoting biases toward humans altogether if our CUIs aren't imagined as human at all [27, 15, 4, 5]. Yet, sidestepping the problem does not contribute to positive social change, meaning our hope to challenge stereotypes and biases in human-human interaction by directly challenging those biases in HCI would be unattainable.

Conclusion

Researchers and designers think and develop tomorrow's conversational user interfaces. If we are aspiring to work in a future-oriented fashion, we need to be as aware of the diverse society we cater to as possible. This means challenging ourselves in pursuing research questions around diverse CUIs as well as challenging users to adopt a more diverse set of CUI personalities. It is not enough to merely identify problems of bias in social representations of today's CUI voices and personas, we must work to actively challenge these biases in our designs and in our research. Doing so may mean subverting norms of both user-centred design and of research design, as challenging entrenched biases may mean limiting the choices of the people who interact with CUIs or choosing CUIs for our laboratory research that differ in design from the commercially popular ones - studying CUIs not as they are, but as they ought to be. Nevertheless, it is our belief that our ethical responsibility is not to uphold and reinforce norms when those norms are exclusionary and perpetuate inequality. It is time that the CUI community designs for the future it wants to see, rather than fitting design into the status quo of the present. Accomplishing this challenges us to design for the users we want, not the users we have.

REFERENCES

[1] Gavin Abercrombie, Amanda Cercas Curry, Mugdha Pandya, and Verena Rieser. 2021. Alexa, Google, Siri: What are Your Pronouns? Gender and Anthropomorphism in the Design and Perception of Conversational Assistants. In *Proceedings of the 3rd Workshop on Gender Bias in Natural Language Processing*. Association for Computational Linguistics, Online, 24–33. DOI: http://dx.doi.org/10.18653/v1/2021.gebnlp-1.4

[2] Aylin Caliskan, Joanna J. Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (2017), 183–186. DOI: http://dx.doi.org/10.1126/science.aal4230

[3] Julie Carpenter. 2019. Why project Q is more than the world's first nonbinary voice for technology. *Interactions* 26, 6 (Oct. 2019), 56–59. DOI: http://dx.doi.org/10.1145/3358912

[4] Philip R Doyle, Leigh Clark, and Benjamin R. Cowan. 2021. What Do We See in Them? Identifying Dimensions of Partner Models for Speech Interfaces Using a Psycholexical Approach. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–14. DOI: http://dx.doi.org/10.1145/3411764.3445206

[5] Philip R. Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R. Cowan. 2019. Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services - MobileHCI '19*. ACM Press, Taipei, Taiwan, 1–12. DOI: http://dx.doi.org/10.1145/3338286.3340116

[6] Justin Edwards, Christian Janssen, Sandy Gould, and Benjamin R. Cowan. 2021. Eliciting Spoken Interruptions to Inform Proactive Speech Agent Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–12. DOI: http://dx.doi.org/10.1145/3469595.3469618

[7] Justin Edwards, He Liu, Zhou Tianyu, Gould Gould, Sandy J. J., Leigh Clark, Philip Doyle, and Benjamin R Cowan. 2019. Multitasking with Alexa: How Using Intelligent Personal Assistants Impacts Language-based Primary Task Performance. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. Dublin, Ireland. Accetped.

[8] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2020. Gender Bias in Chatbot Design. In *Chatbot Research and Design*, Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie Lai-Chong Law, Ole-Christoffer Granmo, Ewa Luger, and Petter Bae Brandtzaeg (Eds.). Vol. 11970. Springer International Publishing, Cham, 79–93. DOI: http://dx.doi.org/10.1007/978-3-030-39540-7_6 Series Title: Lecture Notes in Computer Science.

[9] Leah Fessler. 2017. We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz* (Feb. 2017).

[10] Megan Garcia. 2016. Racist in the Machine. *World Policy Journal* 33, 4 (2016), 111–117. Publisher: JSTOR.

[11] Layne Hubbard, Shanli Ding, Vananh Le, Pilyoung Kim, and Tom Yeh. 2021. Voice Design to Support Young Children's Agency in Child-Agent Interaction. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–10. DOI: http://dx.doi.org/10.1145/3469595.3469604

[12] Gilhwan Hwang, Jeewon Lee, Cindy Yoonjung Oh, and Joonhwan Lee. 2019. It Sounds Like A Woman: Exploring Gender Stereotypes in South Korean Voice Assistants. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–6. DOI: http://dx.doi.org/10.1145/3290607.3312915

[13] Capgemini Research Institute. 2019. *Voice on the go. How can auto manufacturers provide a superior in-car voice experience.* Technical Report. Capgemini Research Institute. CapgeminiResearchInstitute

[14] Calvin K Lai and Mahzarin R Banaji. 2019. The psychology of implicit intergroup bias and the prospect of change. (2019).

[15] Sébastien Le Maguer and Benjamin R. Cowan. 2021. Synthesizing a human-like voice is the easy way. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–3. DOI: http://dx.doi.org/10.1145/3469595.3469614

[16] Peter M. Lewis. 2020. "A Claim to Be Heard": Voices of Ordinary People in BBC Radio Features. *Revue française de civilisation britannique* XXVI, 1 (Dec. 2020). DOI:http://dx.doi.org/10.4000/rfcb.7681

[17] Irene Lopatovska, Elena Korshakova, and Diedre Brown. 2021. Effects of Gendered Voices on Personality Perceptions of Conversational User Interfaces.

[18] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16*. ACM Press, Santa Clara, California, USA, 5286–5297. DOI: http://dx.doi.org/10.1145/2858036.2858288

[19] Donald McMillan and Razan Jaber. 2021. Leaving the Butler Behind: The Future of Role Reproduction in CUI. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*. ACM, Bilbao (online) Spain, 1–4. DOI:http://dx.doi.org/10.1145/3469595.3469606

[20] Roger K Moore. 2017. Appropriate Voices for Artefacts: Some Key Insights. *1st Int. Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots* (2017), 5.

[21] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (Jan. 2000), 81–103. DOI:http://dx.doi.org/10.1111/0022-4537.00153

[22] Donald A. Norman. 2013. *The design of everyday things* (rev. and expanded ed ed.). MIT Press, Cambridge, Mass.

[23] V. Phoebe Pawlik. 2022. Design Matters! How Visual Gendered Anthropomorphic Design Cues Moderate the Determinants of the Behavioral Intention Towards Using Chatbots. In *Chatbot Research and Design*, Asbjørn Følstad, Theo Araujo, Symeon Papadopoulos, Effie L.-C. Law, Ewa Luger, Morten Goodwin, and Petter Bae Brandtzaeg (Eds.). Vol. 13171. Springer International Publishing, Cham, 192–208. DOI: http://dx.doi.org/10.1007/978-3-030-94890-0_12 Series Title: Lecture Notes in Computer Science.

[24] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. ACM Press, Montreal QC, Canada, 1–12. DOI: http://dx.doi.org/10.1145/3173574.3174214

[25] Daniel J. Rea, Yan Wang, and James E. Young. 2015. Check Your Stereotypes at the Door: An Analysis of Gender Typecasts in Social Human-Robot Interaction. In *Social Robotics*, Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi (Eds.). Vol. 9388. Springer International Publishing, Cham, 554–563. DOI: http://dx.doi.org/10.1007/978-3-319-25554-5_55 Series Title: Lecture Notes in Computer Science.

[26] Elayne Ruane, Abeba Birhane, and Anthony Ventresque. 2019. Conversational AI: Social and Ethical Considerations.. In *AICS*. 104–115.

[27] Sergio Sayago and Josep Blat. 2020. Reimagining Communication With Conversational User Interfaces. In *Reimagining Communication: Mediation* (1 ed.), Michael Filimowicz and Veronika Tzankova (Eds.). Routledge, Abingdon, Oxon ; New York, NY : Routledge, 2020., 287–302. DOI: http://dx.doi.org/10.4324/9781351015431-18

[28] Kristen M. Scott, Simone Ashby, and Julian Hanna. 2020. "Human, All Too Human": NOAA Weather Radio and the Emotional Impact of Synthetic Voices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–9. DOI: http://dx.doi.org/10.1145/3313831.3376338

[29] Rob Semmens, Nikolas Martelaro, Pushyami Kaveti, Simon Stent, and Wendy Ju. 2019. Is Now A Good Time?: An Empirical Study of Vehicle-Driver Communication Timing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. ACM Press, Glasgow, Scotland Uk, 1–12. DOI: http://dx.doi.org/10.1145/3290605.3300867

[30] Ben Shneiderman and Catherine Plaisant. 2004. *Designing the user interface: strategies for effective human-computer interaction* (4th ed ed.). Pearson/Addison Wesley, Boston.

[31] Jeff Sugarman. 2015. Neoliberalism and psychological ethics. *Journal of Theoretical and Philosophical Psychology* 35, 2 (May 2015), 103–116. DOI: http://dx.doi.org/10.1037/a0038960

[32] Selina Jeanne Sutton. 2020. Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Bilbao Spain, 1–8. DOI: http://dx.doi.org/10.1145/3405755.3406123

[33] Thomas Teo. 2008. From Speculation to Epistemological Violence in Psychology: A Critical-Hermeneutic Reconstruction. *Theory & Psychology* 18, 1 (Feb. 2008), 47–67. DOI: http://dx.doi.org/10.1177/0959354307086922

[34] UNESCO. 2019. *I'd blush if I could – closing gender divides in digital skills through education.* Technical Report. UNESCO.

[35] Katja Wagner, Frederic Nimmermann, and Hanna Schramm-Klein. 2019. Is It Human? The Role of Anthropomorphism as a Driver for the Successful Acceptance of Digital Voice Assistants. In *HICSS*.

[36] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology* 52 (May 2014), 113–117. DOI: http://dx.doi.org/10.1016/j.jesp.2014.01.005

[37] Yunhan Wu, Justin Edwards, Orla Cooney, Anna Bleakley, Philip R. Doyle, Leigh Clark, Daniel Rough, and Benjamin R. Cowan. 2020a. Mental Workload and Language Production in Non-Native Speaker IPA Interaction. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, Bilbao Spain, 1–8. DOI: http://dx.doi.org/10.1145/3405755.3406118

[38] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R. Doyle, Leigh Clark, and Benjamin R. Cowan. 2020b. See What I'm Saying? Comparing Intelligent Personal Assistant Use for Native and Non-Native Language Speakers. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Oldenburg Germany, 1–9. DOI: http://dx.doi.org/10.1145/3379503.3403563