An Empathy Imitation Game: Empathy Turing Test for Care- and Chat-bots

Jeremy Howick[1] · Jessica Morley[2,3] Luciano Floridi[1,2,3]

Asimo is a robot that helps people with disabilities feed themselves (Honda Motor Company 2007). Robear reduces the physical strain of carers' work by lifting and moving heavy patients (Szondy 2015). Pepper organises sing-songs, makes gin and tonics, and can mirror your tone of voice when speaking to you (Schussler et al. 2020). Carebots have "cousins" called chatbots that simulate human speech but not physical behaviour. They can allegedly provide cognitive behavioural therapy for an array of psychiatric disorders (Palanica et al. 2019). A chatbot called Woebot uses artificial intelligence (AI) to track moods and offers tips it claims can reduce depression, anxiety, and other psychological problems (Fitzpatrick et al. 2017). A chatbot called Shim successfully promoted mental wellbeing using cognitive behavioural therapy (Ly et al. 2017). Carebots and chatbots (henceforth 'artificial carers') are often promoted as a means of coping with the problem posed by a shortage of human carers and a growing elderly population (Pepito and Locsin 2019). Faith in the success of artificial carers assumes that automation-derived efficiency gains in the health and care sectors is likely to mirror those seen in other sectors (Susskind and Susskind 2015).

Tools and technology have assisted doctors for millennia, and computers have improved doctors' diagnoses for decades (Grove et al. 2000). Yet, one may claim that systems like Asimo, Robear, Woebot do more than help. Artificial carers are getting better and better at doing what humans can do at breakneck speed (Floridi et al. 2018), which leads some to surmise that they might actually replace (many functions of) human doctors within a few decades (Goldhahn et al. 2018).

Besides the lack of robust evidence about safety and efficacy, another barrier to this hypothesised future becoming reality, is that it is hard to imagine these artificial carers empathizing, understanding, or offering the same degree of

✉ Jeremy Howick
jeremy.howick@philosophy.ox.ac.uk

[1] Oxford Empathy Programme, Faculty of Philosophy, University of Oxford, Radcliffe Humanities Quarter, Oxford OX2 6GG, UK

[2] Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK

[3] British Library, The Alan Turing Institute, 96 Euston Rd, London NW1 2DB, UK

compassion to patients that warm-blooded, living and breathing humans do. In what follows, we shall focus only on empathy, but parallel arguments apply to compassion, understanding and other abilities and interactions considered to be 'human'. If artificial carers cannot empathize, they are at a great disadvantage because people like empathic practitioners and empathic care improves health outcomes (Howick et al. 2018a, b).

Some may claim that artificial carers cannot empathize *a priori* because, they say, empathy requires a human mind (Fernandez and Zahavi 2020). This philosophical position is an offshoot of more fundamental debates about whatit means to be human or have a human mind. We propose to move this debate from the abstract to the concrete. Taking our inspiration from the Turing Test for human thinking (Turing 1950), we propose to replace "can artificial carers be empathic?*"* with "can a human user distinguish between the empathy showed by an artificial carer and that showed by a human practitioner?". The new question could be answered by asking patients to compare the empathy they *perceive* from an artificial carer with empathy they perceive from a human carer.

The most widely used and best-validated scale used by patients to measure practitioner empathy is the Consultation and Relational Empathy Measure (CARE) (Mercer et al. 2004). The CARE measure asks recipients of care various questions about their consultation, such as whether the practitioner treated them as a whole person, whether they felt understood, and whether the practitioneroffered hope. The CARE measure could be given to recipients of healthcare after a consultation with a human doctor, and after a consultation with an artificial carer. If the CARE scores were similar, then we could tentatively conclude that the artificial carer was (perceived to be) as good at empathizing as the human carer. If the artificial carer's CARE score was lower, then we could conclude that the artificial carer fell short when it came to empathy.

To make our Turing-type test fair, some small modifications to CARE might be required so that the language and concepts are agnostic with respect to the human or artificial nature of the "practitioner". For example, question (5) in the CARE form asks to evaluate "How good was the practitioner at fully understanding your concerns?" A patient aware that the carer is artificial (Asimo and friends are obviously robots) may find it less confusing to be asked a question that involved referring to a robot as a practitioner. There are, of course, critics of CARE (Hong and Han 2020), and these critics could choose their own method for comparing the empathy of the practitioners with that of the artificial carers.

Our proposed approach has a number of advantages as well as some pitfalls. An advantage is that it sidesteps unresolved debates about the nature of human minds, and the definition of empathy (Howick et al. 2018a, b), and gets straight to what matters to patients. A key question to individuals receiving care is whether the carer (human or bot) behave in a way that the individual feels empathised with, enabling them to feel better, and recover faster. From this perspective, the empathic inner state of the empathizer is more important than the behaviours which lead someone to feel empathized with (Howick et al. 2018a, b).

A pitfall of our suggestion is that it falls prey to all the objections to Turing-like tests, especially the *in principle* impossibility of detecting consciousness (which

applies if 'true' empathy requires a conscious being) (Oppy and Dowe 2020). To these we can add a number of potential ethical and legal problems. These include possible deception (is the artificial carer 'pretending' to care?), infringement of privacy (where is the artificial carer storing data?) (Floridi and Cowls 2019), and the proliferation of subtle bias. Artificial carers certainly should—and probably can—be designed to avoid these ethical pitfalls (Yew 2020).

Also, no matter how safe, effective, and empathic artificial carers might be, it does not follow that we should use them. Their use affects a number of dimensions, ranging from disruptions to the workforce to the inability to make value-based ethical decisions (although the latter could also be subject to a related Turing-type test) (Powell 2019). These challenges make it more likely that artificial carers will complement rather than replace most human functions. For example, they might take over many of the menial tasks human carers currently do, and thus free up time for doctors to offer empathy. Somewhat counterintuitively, this would lead to a situation whereby "AI will allow doctors to be more human" (Academy of Medical Royal College 2019).

The groundwork for our proposal is already being laid. Thirty-one children receiving IV placement were able to distinguish between robots programmed to be empathic from robots not programmed to be empathic (Trost et al. 2020); brain scans have shown that humans perceive empathy from robots (Suzuki et al. 2015); and people asked to rate *descriptions* of robots were able to distinguish between empathic and less empathic robots (Chita-Tegmark et al. 2019).

We do not believe that care- or chat-bots will outperform very empathic human clinicians in our proposed test. However, the extent to which people feel that their practitioners are empathic varies from exceptional to wanting (Howick et al. 2017). As it is happening in other professions (Susskind and Susskind 2015), it is reasonable to assume that the best artificial carers may outperform the worst human practitioners, especially as technology progresses. And, from a person-centred perspective, being treated kindly may take precedence over the source of kindness (human or artificial) or their motivations (intentional or not). Moreover, in light of the growing need for human practitioners, we believe that patients will benefit when philosophical debates about the extent to which artificial carers can be empathic are sidestepped in favour of rigorous Turing-type tests that compare perceived empathy of a care or chatbot with perceived empathy of a human practitioner. This will pave the way for artificial carers to replace many of the repetitive, administrative, and menial tasks that currently stand in the way of human carers offering empathy. Our proposed tests, and any inferences drawn from these tests' results, need to be deployed in the context of a robust and ethical framework for the deployment of effective and safe artificial carers that fully respect of human values and dignity.

References

Academy of Medical Royal College. (2019). Artificial intelligence in healthcare. Retrieved from https://www.aomrc.org.uk/reports-guidance/artificial-intelligence-in-healthcare/. Accessed 8 Dec 2020

Chita-Tegmark, M., Ackerman, J. M., & Scheutz, M. (2019). Effects of assistive robot behavior on impressions of patient psychological attributes: Vignette-based human-robot interaction study. *Journal of Medical Internet Research, 21*(6), e13729.

Fernandez, A. V., & Zahavi, D. (2020). Basic empathy: Developing the concept of empathy from the ground up. *International Journal of Nursing Studies, 110,* 103695.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4*(2), e19.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review*. https://doi.org/10.1162/99608f92.8cd550d1.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., et al. (2018). AI4Peo- ple- an ethical framework for a good AI society: Opportunities, risks, principles, and recommen- dations. *Minds and Machines, 28*(4), 689–707.

Goldhahn, J., Rampton, V., & Spinas, G. A. (2018). Could artificial intelligence make doctors obsolete? *BMJ, 363,* k4563.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment, 12*(1), 19–30.

Honda Motor Company. (2007). The ASIMO technical robotics manual: ASIMO history and specifications. Retrieved from http://asimo.honda.com/Abstract-Technical-Information. Accessed 30 Sept 2020.

Hong, H., & Han, A. (2020). A systematic review on empathy measurement tools for care professionals. *Educational Gerontology, 46*(2), 72–83.

Howick, J., Bizzari, V., Dambha-Miller, H., & Oxford Empathy, P. (2018a). Therapeutic empathy: What it is and what it isn't. *Journal of the Royal Society of Medicine, 111*(7), 233–236.

Howick, J., Moscrop, A., Mebius, A., Fanshawe, T. R., Lewith, G., Bishop, F. L., et al. (2018b). Effects of empathic and positive communication in healthcare consultations: A systematic review and meta-analysis. *Journal of the Royal Society of Medicine*. https://doi.org/10.1177/0141076818769477.

Howick, J., Steinkopf, L., Ulyte, A., Roberts, N., & Meissner, K. (2017). How empathic is yourhealthcare practitioner? A systematic review and meta-analysis of patient surveys. *BMC Medical Education, 17*(1), 136.

Ly, K. H., Ly, A. M., & Andersson, G. (2017). A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. *Internet Interventions, 10,* 39–46.

Mercer, S. W., Maxwell, M., Heaney, D., & Watt, G. C. (2004). The consultation and relational empa- thy (CARE) measure: Development and preliminary validation and reliability of an empathy-based consultation process measure. *Family Practice, 21*(6), 699–705.

Oppy, G., & Dowe, D. (2020). The turing test. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy*. Stanford: Stanford University.

Palanica, A., Flaschner, P., Thommandram, A., Li, M., & Fossat, Y. (2019). Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *Journal of Medical Internet Research, 21*(4), e12887.

Pepito, J. A., & Locsin, R. (2019). Can nurses remain relevant in a technologically advanced future? *International Journal of Nursing Sciences, 6*(1), 106–110.

Powell, J. (2019). Trust me, I'm a chatbot: How artificial intelligence in health care fails the turing test. *Journal of Medical Internet Research, 21*(10), e16222.

Schussler, S., Zuschnegg, J., Paletta, L., Fellner, M., Lodron, G., Steiner, J., et al. (2020). Effects of a humanoid socially assistive robot versus tablet training on psychosocial and physical outcomes of persons with dementia: Protocol for a mixed methods study. *JMIR Research Protocols, 9*(2), e14927.

Susskind, R., & Susskind, D. (2015). *The future of the professions: How technology will transform the work of human experts*. Oxford: Oxford University Press.

Suzuki, Y., Galli, L., Ikeda, A., Itakura, S., & Kitazaki, M. (2015). Measuring empathy for human and robot hand pain using electroencephalography. *Scientific Reports, 5,* 15924.

Szondy, D. (2015). Robear robot care bear designed to serve Japan's aging population. Retrieved from https://newatlas.com/robear-riken/36219/. Accessed 30 Sept 2020

Trost, M. J., Chrysilla, G., Gold, J. I., & Mataric, M. (2020). Socially-Assistive robots using empathy to reduce pain and distress during peripheral IV placement in children. *Pain Research and Management, 2020,* 7935215.

Turing, A. (1950). Computing machinery and intelligence. *Mind, LIX*(236), 433–460.

Yew, G. C. K. (2020). Trust in and ethical design of carebots: The case for ethics of care. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-020-00653-w.