HUMAN-ROBOT INTERACTION

Social robots in rehabilitation: A question of trust

Philipp Kellmeyer[1,2,3], Oliver Mueller[2,4], Ronit Feingold-Polak[5], Shelly Levy-Tzedek[5,6]*

Social robots can help meet the growing need for rehabilitation assistance; measures for creating and maintaining trust in human-robot interactions should be priorities when designing social robots for rehabilitation.

The need for effective, scalable rehabilitation strategies is projected to grow substantially in the coming decades, with increased patient survival after diseases with severe functional deficits, such as stroke. Socially assistive robots (SARs) use voices, gestures, or other human-like behaviors for social interaction or functional assistance (Fig. 1) (*1*). Developing SARs to assist in rehabilitation is an important emerging research area of medical technology.

In rehabilitation, SARs engage with persons from an especially vulnerable population. Patients depend on reliable and effective relationships with their therapists. They are often older and multimorbid and suffer from psychological distress because of their disability (*2*). Complex injuries—which may entail motor, cognitive, speech, and language impairments—call for a comprehensive interaction design that addresses these aspects. Although deployment of SARs may support rehabilitation in a pioneering way, it raises important ethical and societal questions.

We argue that trust is a central condition for successful interactions between patients and SARs in a rehabilitation context (*3*). We propose that safety, shared intentionality, predictability of behavior, and mutual attunement are particularly salient prerequisites for establishing trust between humans and SARs. Whereas, in philosophy, a basis for a trusting relationship is the ability to infer another's internal state from behavior, in medicine, acknowledging the patients' vulnerabilities is key. In engineering, enabling robots that capture, interpret, and predict human movements is a way to enhance trust.

(1) Safety of human-robot interaction. The patient must be confident that the SAR performs movements that promote and support therapy and does no harm (*4*). In stroke rehabilitation, repetition is key, but repeating incorrect movements can be detrimental to the process. In industry, engineers aim to design robots with maximum predictability of the system's behavior. Such rigidity might be less appropriate for close human-robot interaction in medicine, because patients might quickly lose interest in working with the robot. However, building more "intelligence," or autonomy, into medical robots might diminish their predictability and, consequently, patient safety. A safe SAR is certainly a conditio sine qua non, but it may not suffice. The challenge is to design SARs that satisfy accepted safety requirements yet are enticing and motivational for humans to interact with.

(2) Shared intentionality and predictability of actions. For SARs to be accepted as partners during rehabilitation, patients should be able to recognize and predict the SAR's intentions and respective gestures (*5*) and facial expressions. Approaches combining philosophy and cognitive science—particularly the 4E framework of embodied, embedded, extended, and enactive cognition (*6*)—could help to analyze emerging types of human-robot interactions (*7*). Studies in developmental psychology suggest that sharing intentionality is the basis for successful interactions (*8*). Although robots do not have their own intentions, it is useful to apply theories on joint agency and shared intentions to robotics, because the trust in these devices may increase if patients feel that the SAR shares their goals and intentions. Patients with cognitive impairment, however, may experience difficulty understanding the intentions of the robot, and thus an adaptive design should be able to adjust to different levels of cognitive ability. Designing SAR as a trustworthy partner also touches the issues of intersubjectivity and relational agency, and deceit of the patient (e.g., the "Wizard of Oz" approach, where, unbeknown to participants, another person is operating the robot remotely) should be considered only when the benefits outweigh the risk of breaking the trust that the human has developed in the robot.

(3) Mutual attunement between SAR and human. SARs in rehabilitation often take the form of exercise coaches (*1*). Because performance by patients varies considerably (due to, e.g., fatigue or reduced motivation), the degree to which the robot attunes to the emotional, psychological, and physical condition of the patient influences the success of the interaction. SARs offer the potential for verbal communication with the patient during exercise. However, patients often suffer from some speech impairment after stroke. This can compromise the ability of the robot to understand the patient's verbal communication. Such problems may frustrate the patient and diminish trust in the robotic partner. Similarly, the degree to which the patient can attune to the robot's range of behaviors, its variability in performance (including errors), and its human likeness will also determine the success of robot-assisted therapy.

The patient's personal history (e.g., technology acceptance) will also likely affect the degree of trust he or she feels vis-à-vis the robot (*9*) as well as the functional role he or she assigns to the robot (e.g., partner, tool, coach). In Fogg's "functional triad" (*10*), for technology to be persuasive, humans need to recognize it not only as a tool but also as a medium and a social actor.

Considering these prerequisites for trusting human-robot relationships, along with a comprehensive study of users' preferences (in terms of proxemics, kinematics, etc.) regarding the robot's behavior and degrees of autonomy, could make SARs more effective for rehabilitation. However, it is an open empirical question whether adaptive SARs perform better in augmenting rehabilitation treatments than robotic orthoses for supporting movements, when combined with gamified simulations.

With the increasing use of SARs in medicine, the question of trust may extend from

[1]Translational Neurotechnology Lab, Department of Neurosurgery, University of Freiburg–Medical Center, Freiburg im Breisgau, Germany. [2]Cluster of Excellence BrainLinks-BrainTools, University of Freiburg, Germany. [3]Institute for Biomedical Ethics and History of Medicine, University of Zurich, Switzerland. [4]Department of Philosophy, University of Freiburg, Germany. [5]Recanati School for Community Health Professions, Department of Physical Therapy, Ben-Gurion University of the Negev, Beer-Sheva, Israel. [6]Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
*Corresponding author. Email: shelly@bgu.ac.il

the individual patient interacting with a robot to the health care system at large. In our opinion, SARs should not replace human physical therapists but rather augment the range of therapeutic options. SARs may be better suited for certain tasks (e.g., performing repetitive operations without fatiguing or becoming impatient), but the depth of the interpersonal interaction and the knowledge that humans bring to the therapy are invaluable. Thus, implemented alongside a therapist, SARs may open up richer opportunities for exercises in rehabilitation.

It is crucial that research on the integration of SARs in rehabilitation encompasses a user-centered interdisciplinary approach that considers the issues raised here and integrates the perspectives of the patient, the health care system, and our society. The successful implementation of SARs in rehabilitation will depend not only on technological improvements but also on the integration of SARs as members in a social environment. Building trust between humans and SARs should lie at the heart of this integration process.

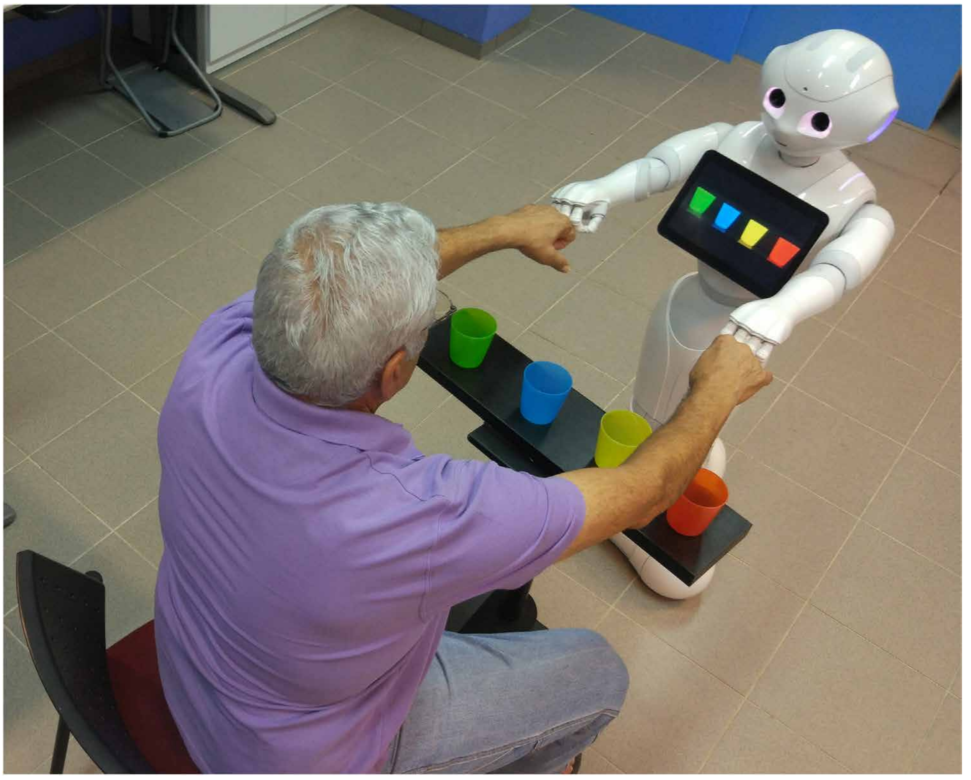

Fig. 1. An example of an interaction with a SAR. The robot Pepper (Softbank Robotics) and a participant successfully exchange a social gesture ("fist bump") after the participant arranged a set of colored cups according to an image on Pepper's screen.

REFERENCES AND NOTES

1. M. J. Matarić, Socially assistive robotics: Human augmentation versus automation. *Sci. Robot.* **2**, eaam5410 (2017).
2. S. Mason, J. Wardrope, G. Turpin, A. Rowlands, The psychological burden of injury: An 18 month prospective cohort study. *Emerg. Med. J.* **19**, 400–404 (2002).
3. P. Lin, K. Abney, R. Jenkins, *Robot Ethics 2. 0: New Challenges in Philosophy, Law, and Society* (Oxford Univ. Press, 2017).
4. M. Iosa, G. Morone, A. Cherubini, S. Paolucci, The three laws of neurorobotics: A review on what neurorehabilitation robots should do for patients and clinicians. *J. Med. Biol. Eng.* **36**, 1–11 (2016).
5. A. Sciutti, M. Mara, V. Tagliasco, G. Sandini, Humanizing human-robot interaction: On the importance of mutual understanding. *IEEE Technol. Soc. Mag.* **37**, 22–29 (2018).
6. R. Menary, Introduction to the special issue on 4E cognition. *Phenom. Cogn. Sci.* **9**, 459–463 (2010).
7. A. Lavazza, Freedom of thought and mental integrity: The moral requirements for any neural prosthesis. *Front. Neurosci.* **12**, 82 (2018).
8. M. Tomasello, M. Carpenter, Shared intentionality. *Dev. Sci.* **10**, 121–125 (2007).
9. I. Gaudiello, E. Zibetti, S. Lefort, M. Chetouani, S. Ivaldi, Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Comput. Human Behav.* **61**, 633–655 (2016).
10. B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do* (Morgan Kaufmann, 2003).

Acknowledgments: We thank two anonymous reviewers for helpful comments that improved the manuscript. P.K., O.M., and S.L. will be Research Fellows during the academic year 2018 -2019 at the Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Germany. **Funding:** This work was partly supported by the German Ministry of Education and Research (BMBF) grant 13GW0053D (MOTOR-BIC) to the University of Freiburg–Medical Center and the German Research Foundation (DFG) grant EXC 1086 BrainLinks-BrainTools to the University of Freiburg, Germany. The research was also partly supported by the Helmsley Charitable Trust through the Agricultural, Biological and Cognitive Robotics Initiative and by the Marcus Endowment Fund, both at the Ben-Gurion University of the Negev. Further financial support was provided by the Promobilia Foundation and the Borten Family Foundation grants, by the Israel Science Foundation (grant nos. 535/16 and 2166/16), and by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 754340. **Competing interests:** The authors declare that they have no competing interests.

10.1126/scirobotics.aat1587

Citation: P. Kellmeyer, O. Mueller, R. Feingold-Polak, S. Levy-Tzedek, Social robots in rehabilitation: A question of trust. *Sci. Robot.* **3**, eaat1587 (2018).

Science Robotics

Social robots in rehabilitation: A question of trust

Philipp Kellmeyer, Oliver Mueller, Ronit Feingold-Polak and Shelly Levy-Tzedek

Sci. Robotics **3**, eaat1587.
DOI: 10.1126/scirobotics.aat1587