Title Teaching robots the art of human social synchrony

Rachael E. Jack*
School of Psychology & Neuroscience, University of Glasgow, Scotland, UK

*Corresponding author rachael.jack@glasgow.ac.uk

Abstract Humanoid robots can now learn the art of social synchrony using neural networks.

Main text

In the digital age, social robots are fast becoming part of mainstream society, from training doctors and educating children, to providing talking therapies and customer service. Now a global multi-billion-dollar industry, the increasing demand for robots with human-like social intelligence marks a significant milestone in our technological history. Thanks to rapid developments in artificial intelligence (AI), robots—once primarily confined to dull, dirty and dangerous work, such as stocking shelves, cleaning floors, deactivating bombs—are now elevated to join the human social world, with immense transformative potential for society(*1*). Furthermore, as Human Digital Twins, such robots can serve as invaluable tools for scientific inquiry, enabling researchers to simulate, study, and better understand complex human social behaviors and cognitive processes(*2*).

However, if you have ever interacted with a social robot, you would quickly realize the limits of this potential. Although their physical appearance is increasingly impressive, their interactions are often clunky, stilted and awkward. Something feels off. Why? One critical limitation is that current social robots lack the art of social synchrony, where nods, smiles, gestures and speech are carefully orchestrated across conversation partners(*3*). Though seemingly effortless for most of us, such turn-taking is a highly complex skill that involves rapidly processing others' speech, vocal tone, facial expressions and gestures, and planning precisely when and how to respond(*4*). Such sophisticated exchange dynamics are present in every language, spoken and signed, and is widely considered to be the fundamental 'engine' for successful social interaction(*5*). Therefore, for social robots to engage in human social interactions, social synchrony skills are essential. Yet, as with many other human social behaviors, equipping social robots with a sophisticated human-like social intelligence is challenging(*6*).

In their recent paper, Hu *et al.(7)* addressed the art of social synchrony by endowing one such humanoid social robot, EMO, with it. EMO is a soft-skinned anthropomorphic facial robot that can display a wide range of nuanced facial expressions using 26 magnet-controlled facial actuators. It also has high-resolution cameras in its eye sockets to detect different types of facial expressions. Although EMO can mimic the human facial expressions it detects, engaging in social synchrony involves a more refined planning and execution of responses. To achieve this, Hu and colleagues trained EMO using neural networks to predict the facial expressions of its human interlocutors based on their early facial movements. For example, detecting raised corners of the mouth suggests that a happy facial expression will be displayed. EMO's new predictive ability enables it to plan and execute its own facial expression in response, achieving a more human-like social synchrony. Hu and colleagues also upgraded EMO processing capacity to run on lightweight computing facilities, freeing up processing power for the development of other functions, such as speech and listening.

Using this simple and elegant approach, Hu and colleagues elevated EMO's social interaction skills from mere mimicry (Figure 1A) to the art of social synchrony (Figure 1B). Such a development has profound implications for the future of social robots. For example, even in this nascent form based solely on facial expressions, these turn-taking skills could radically improve trust and rapport building in human-robot interactions, bringing social robots one step closer to reaching their potential. Importantly, the success of EMO's new skills will depend on what its human conversation partners make of them. Do they improve social exchanges with its human users, engendering trust and empathy? Or is something still off? If so, what is it? Given that EMO's facial expressions can be precisely controlled, a fruitful approach would be to use data-driven social psychophysics methods from the human behavioral sciences(*8*). Specifically, such methods would experimentally manipulate different features of EMO's expressions, such as the exact timing of when they start or their specific facial movement components, and test how they influence human user behaviors, such as how much they trust or engage with EMO. By pin-pointing the specific features that facilitate or hinder human user engagement, such feedback could then be used to improve EMO's social communication skills for both general use and bespoke applications, including cross-cultural interactions(*9*), thereby enhancing its utility, accessibility and marketability(*10*).

Finally, human social interactions are inherently multi-modal, involving complex combinations of visual and auditory signals, such as nodding, "uhms" and "ahhs," raised eyebrows, averted gaze, long blinks, and hand gestures. Here, Hu and colleagues focus on a single modality—facial expressions—but their results pave the way to developing social synchrony skills with more complex multi-modal signals. Such a feat, though a complex interdisciplinary endeavour, could truly enable social robots to join the human social world.

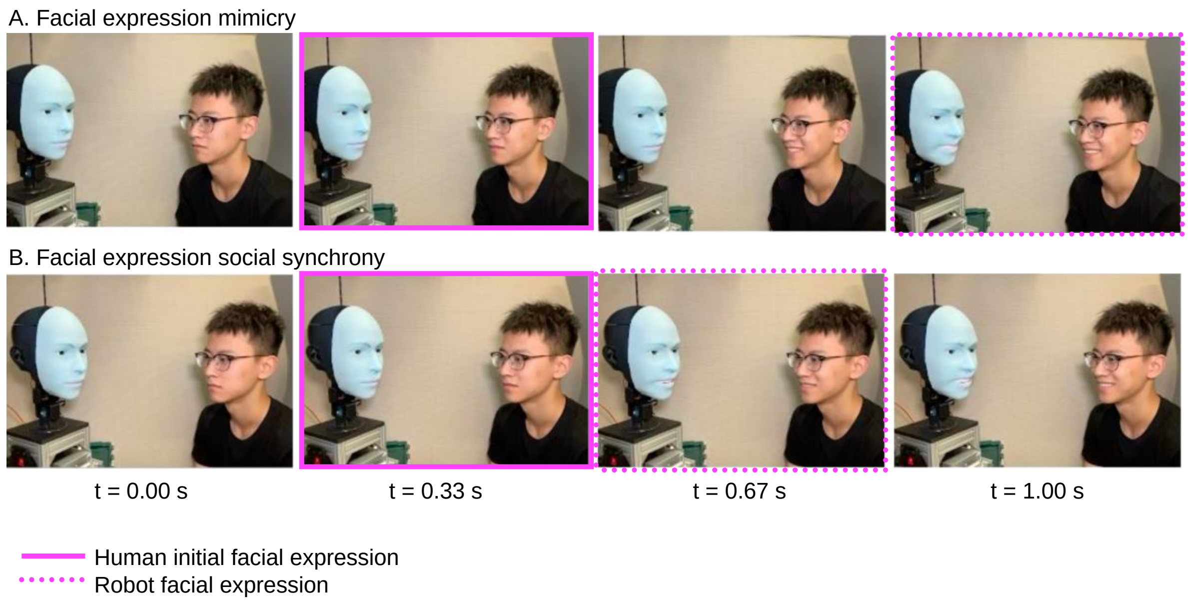

Figure 1. Facial expression mimicry vs facial expression social synchrony in robot-human interaction. A. In facial expression mimicry, the social robot copies the human's facial expression which inevitably involves displaying it with a noticeable delay. Note for example, at time point 3, the mismatch in facial expression between the robot and human. **B**. In facial expression social synchrony, the robot can predict the human's facial expression from early facial movements (see time point 2). This enables the robot to synchronize its facial expressions with the human's. Note for example, at time point 3, that the robot and

human each display the same facial expression at the same time. Figure Adapted from Hu et al (2024).

References

1. B. Lugrin, C. Pelachaud, D. Traum, *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*. (ACM, 2022).
2. F. Barachini, C. Stary, *From digital twins to digital selves and beyond: Engineering and social models for a trans-humanist world*. (Springer Nature, 2022).
3. S. C. Levinson, Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences* **20**, 6-14 (2016).
4. M. J. Pickering, S. Garrod, The use of prediction to drive alignment in dialogue. *Grounding sociality. Neurons, mind and culture*, 175-191 (2011).
5. S. C. Levinson, in *Roots of human sociality*. (Routledge, 2020), pp. 39-69.
6. C. Breazeal, K. Dautenhahn, T. Kanda, Social robotics. *Springer handbook of robotics*, 1935-1972 (2016).
7. Y. Hu, BoyuanLin, JiongWang, YunzheWang, Yingke Mehlman, CameronLipson, Hod, Human-Robot Facial Co-expression. *Science Robotics* **9**, XXX (2024).
8. R. E. Jack, P. G. Schyns, Toward a social psychophysics of face communication. *Annual Review of Psychology* **68**, 269-297 (2017).
9. C. Chen *et al.*, in *2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019)*. (IEEE, 2019), pp. 1-8.
10. A. Akinade, Y. Haile, N. Mutangana, C. Tucker, D. Vernon, Culturally competent social robots target inclusion in Africa. *Science Robotics* **8**, eadm7012 (2023).