Dances with Robots: Navigating Power Imbalances in Behavioural Signal Exchanges

Kate Hone
Department of Computer Science
Brunel University of London
London UK
kate.hone@brunel.ac.uk

ABSTRACT

This paper examines the role of non-verbal social signals in human–robot interaction, drawing on established findings from human communication research and recent developments in automated social signal processing. I argue that current regulatory approaches, particularly within the EU AI Act, insufficiently address the ways robots may use behavioural signals to influence interaction outcomes.

1 Introduction

The importance of non-verbal signals in human-human interaction is well recognised. This 'second channel' of human communication guides much human interaction, often without being consciously acknowledged [3]. Evidence suggests the combination of behavioural cues or 'social signals' displayed by an individual can produce non-random changes in others [4]. Social signals processing emerged to automatically recognise these signals, initially to develop better understanding of human-human interaction, but increasingly for application within computational systems. Following affective computing as envisioned by Picard [13], researchers argue that giving computers the ability to recognise and respond to social signals should improve human-computer interactions, partly based on the assumption that people automatically treat computers like people [14]. Similar arguments extend to human-robot interaction, with many asserting that for robots to work effectively within human social environments, they must make effective use of social signals. By their nature, social signals involve reciprocal exchange between interaction partners. When one partner is not human, this raises new questions and challenges. This paper introduces the concept of social signals in human interaction and work enabling machine processing of social signals. It then considers implications of robots recognising and displaying social signals, highlighting challenges for effective human-robot interaction and potential ethical and legal implications.

2 Non-verbal signals in human-human interaction

There is longstanding interest in studying non-verbal signals within human-human interaction. Current conceptions move away from classic 'body language' as static codes or universal signals towards understanding in terms of a multimodal, context-bound, culturally shaped, and functionally oriented system [10]. Importantly, signal exchange plays out over time in a reciprocal 'dance' between interlocutors. Automated measurement approaches provide new insights into how multiple channels combine dynamically to regulate outcomes such as turn-taking and trust, highlighting the importance of feedback loops between dialogue partners [10].

For social signals exchange to play a functional role, humans must both perceive (decode) and send (encode) signals. Adams and Kveraga's [1] discussion of how humans process social signals draws on Gibson's theory of direct perception [5] and argues for shared social affordances across various visual cues, suggesting compound social cue processing is characteristic of the human visual system. They argue some processing (and resulting action driven by 'affordance' of the social environment) can occur at a subconscious level (so-called 'bottom-up' processing). While not fully discounting 'top-down' processing and socially learned affordances, this suggests that while humans may naturally detect a range of social signals and spontaneously act upon them, this information may not be available to conscious thought. When asked to make inferences based on others' non-verbal behaviour, people seem moderately good at reflectively interpreting these, though with wide variations in interpersonal perception accuracy, including individual and contextual differences [6].

When considering how humans send (encode) non-verbal signals, we also see two channels. Much signalling is automatic and non-conscious, but people can control elements of gaze, gesture, movement, and prosody to influence understanding [6]. Argyle [2] identifies three reasons why emotional signals are sent: (1) Direct physiological reactions with no communicative intent (e.g., facial disgust when smelling spoiled food); (2) Spontaneous expressions thought to have evolved as social signals (e.g., fear expressions informing others of dangerous stimuli); (3) Deliberately sent emotional expressions not necessarily reflecting experienced emotional state (e.g., smiling when receiving an unwanted present). The third type is most common and depends on the presence of others. Kraut and Johnston [9] found people were 5-10 times more likely to smile at other people than at inanimate stimuli. People also try to conceal or emphasise spontaneous displays depending on social conditions [2]. Signals displayed therefore include deliberate expressions (some genuine, some concealing internal state) and likely include expressions reflecting underlying state which the person cannot consciously control. Other signals regulate interactions (e.g., turn-taking), operating so rapidly they function below conscious awareness.

Pentland [12] argues for evolutionarily developed 'honest signals' including activity level (reflecting interest), influence (how one person's patterns affect another), consistency (showing focus), and mimicry. Collectively these offer insights into intentions, goals, and values; they operate at an unconscious level but predict outcomes in situations such as dates, interviews, and negotiations. Though said to be difficult to fake, work in my own group giving participants feedback on these signals has shown people can improve self-presentation and outcomes in tasks such as media presentation and negotiation [11, 7]. Social signals play out between people over time and cannot be narrowed to separate consideration of 'sender' and 'receiver'.

3 Machine recognition of non-verbal and social signals

There is substantial research aiming to give machines capacity to recognise human non-verbal signals, both to study human-human interaction and to influence human-computer interaction. Many early attempts focused on single channels (gesture, facial expression, tone of voice), but recent state-of-the-art tends towards deep learning with multimodal inputs. Much work is framed around emotion recognition within affective computing, though 'emotion' itself is contested. Social

HRI 2026 Workshop on Empowering Human-Robot Communication Through Non-Linguistic Pathways (HuRoCo @ HRI 2026).

signals processing has focused beyond individuals on signal exchange between interlocutors [16]. Automated systems can potentially detect cues invisible to human dialogue partners, such as autonomic nervous system activity from infrared imaging and skin conductance measures. While accurate emotion categorisation from non-verbal signals in the wild and across cultures remains an open research aim, machines can distinguish some signals beyond human perception and others humans may be unaware they are showing (e.g., pupil dilation is easily measured but people are not conscious of this aspect of their appearance). This creates potential power imbalances between sender and receiver requiring careful design attention.

4 Regulatory context

The European Union's Artificial Intelligence Act (Regulation (EU) 2024/1689) represents an early attempt to mitigate potential harms from AI, including emotion recognition. It regulates emotion recognition systems, defined as AI systems identifying or inferring emotions or intentions of natural persons based on their biometric data. Article 5(1)(f) prohibits deploying AI systems inferring emotions in workplace and educational settings, with narrow exceptions for medical or safety purposes. Emotion recognition systems deployed elsewhere are defined as high risk, requiring providers to meet stringent requirements including informing users they are exposed to emotion recognition systems.

A critical distinction in the EU AI Act concerns systems responding to non-verbal signals without explicitly inferring emotions. Article 3(41) excludes 'the mere detection of readily apparent expressions, gestures or movements unless they are used for identifying or inferring emotions. As an example, detecting a smile is deemed acceptable but using it to infer happiness is not. This suggests systems adapting to behavioural signals could fall outside regulation, as long as signals are not explicitly interpreted as labelled emotions. This distinction creates a grey area where systems may technically comply with the law while leveraging highly sensitive behavioural cues to shape user behaviour.

5 Social signals for human-robot interaction

Several authors argue robots' effective use of non-verbal signals is critical for effective engagement with humans [15]. Considering robots as encoders (displaying non-verbal signals for human users to perceive), various studies have investigated how robots might best use non-verbal signals within human-robot interaction scenarios. An interesting recent perspective from Huang et al. [8] considers how Pentland's 'honest signals' might apply to human-robot interaction, providing appropriate affordances for interaction between what they describe as 'mismatched communication partners'.

Here we consider robots as decoders – what should robots be enabled to recognise from human interlocutors' behaviour? As discussed, EU law brings stringent restrictions when AI processing is deemed to infer emotion from non-verbal behaviour. However, a generative AI system not explicitly labelling 'emotions' could potentially use non-verbal signal recognition to powerful effect to influence interaction outcomes without necessarily falling within the EU regulation's remit. This raises ethical concerns and I argue the legislation's focus is misaligned with the intention behind creating ethical and responsible AI. Distinguishing which signals robots should decode requires considering not only technical feasibility but also user expectations and contextual appropriateness of machine-mediated social meaning. Consider some relevant scenarios:

1. Deliberately sent or exaggerated signals: Because intentionally produced, their recognition aligns with user expectations. Ethically, recognising such signals seems reasonable provided the system does not over-interpret them.

2. Automatic signals used in regulation of dialogue flow: These cues support turn-taking and pacing. Their use in robots appears unproblematic and is unlikely to fall under emotion-recognition restrictions.

3. Automatic 'honest signals': Though not labelled as emotions, these cues predict outcomes in negotiation, persuasion, and other contexts. A system detecting and responding to these could subtly influence user decisions, raising ethical issues not clearly addressed in current regulation.

4. Automatic signals indicating underlying psychological or physiological state: These involuntary cues may expose internal states users neither intend nor expect to share. Because they can imply emotion or intention, their interpretation is likely to fall under the EU AI Act, correctly aligning with its aim to prevent exploitation of sensitive, non-voluntary behavioural information.

The current legal distinction is whether signals are encoded as an emotion. I argue an important additional distinction should be the intended use of information within the interaction's context. Future AI systems will have potential to recognise signals users don't intend to transmit at a level potentially beyond human capacity to recognise and use those signals.

6 Conclusion

Overall, the current regulatory emphasis on whether a system labels behavioural cues as 'emotions' obscures more important questions about how those cues are used within an interaction. Many signals shaping human-human encounters operate below conscious awareness, and their machine-mediated use carries implications for autonomy, persuasion, and fairness deserving greater scrutiny. I argue future regulation should shift from a focus on emotional inference to more nuanced consideration of intended use, contextual expectations, and potential impact on users' decision-making. Such reframing would better support development of socially responsible human-robot interaction.

REFERENCES

[1] Adams, R. B. and Kveraga, K. 2015. Social vision: Functional forecasting and the integration of compound social cues. Review of Philosophy and Psychology 6, 4, 591–610.

[2] Argyle, M. 1988. Bodily communication. Routledge, Abingdon-on-Thames.

[3] Buchanan, M. 2009. Behavioural science: secret signals. Nature 457, 7229, 528–530.

[4] Burgoon, J. K., Magnenat-Thalmann, N., Pantic, M. and Vinciarelli, A. 2017. Social Signals Processing. Cambridge University Press.

[5] Gibson, J.J. 1979. The ecological approach to visual perception. Houghton Mifflin, Boston.

[6] Hall, J., Horgan, T., & Murphy, N. 2019. Nonverbal Communication. Annual Review of Psychology 70, 271-294.

[7] Hooker, N. 2022. Negotiation Skills Training Intervention Based on Automated Recognition of Human Emotion and Nonverbal Behaviour. Thesis, Brunel University of London.

[8] Huang, G., Chen, C., Moore, R.K. 2026. Mismatched Partners in Human-Robot Interaction: Biomimetic Insights on Aligning Internal States with External Expressions. In Living Machines 2025. Lecture Notes in Computer Science, vol 15582. Springer.

[9] Kraut, R.E. and Johnston, R. E. 1979. Social and emotional messaging of smiling: an ethological approach. Journal of Personality and Social Psychology 37, 1539-1553.

[10] Patterson, M., Fridlund, A., & Crivelli, C. 2023. Four Misconceptions About Nonverbal Communication. Perspectives on Psychological Science 18, 1388-1411.

[11] Pereira, M. and Hone. K. 2021. Communication Skills Training Intervention Based on Automated Recognition of Nonverbal Signals. In CHI '21, May 08-13, Yokohama, Japan. ACM.

[12] Pentland, A. 2008. Honest signals: how they shape our world. MIT Press.

[13] Picard, R. 2000. Affective Computing. MIT Press.

[14] Reeves, B. and Nass, C. 1996. The media equation: How people treat computers, television and new media like real people. Cambridge, UK.

[15] Saunderson, S., Nejat, G. 2019. How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human–Robot Interaction. Int J of Soc Robotics 11, 575–608.

[16] Vinciarelli, A., Pantic, M., Heylen, D., Pelachaud, C., Poggi, I., D'Errico, F., et al. 2012. Bridging the gap between social animal and unsocial machine: a survey of social signal processing. IEEE Trans Affective Computing 3, 69–87.