

This article has been published in a revised form in Behavioral and Brain Sciences <http://doi.org/10.1017/S0140525X22001492>. This version is published under a Creative Commons CC-BY-NC-ND licence. No commercial re-distribution or re-use allowed. Derivative works cannot be distributed. © The Author(s), 2023.

Behavioral and Brain Sciences

“Who’s there?”: Depicting Identity in Interaction

--Manuscript Draft--

Manuscript Number:	
Full Title:	“Who’s there?”: Depicting Identity in Interaction
Short Title:	"Who's there?"
Article Type:	Open Peer Commentary
Corresponding Author:	Patrick George, Timothy Healey, Ph.D. Queen Mary University of London London, UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Queen Mary University of London
Corresponding Author's Secondary Institution:	
First Author:	Patrick George, Timothy Healey, Ph.D.
First Author Secondary Information:	
Order of Authors:	Patrick George, Timothy Healey, Ph.D.
	Dr Christine Howes, PhD
	Professor Eleni Gregoromichelaki, PhD
	Dr Greg Mills, PhD
	Professor Matthew Purver
	Dr Arash Eshghi
	Dr Julian Hough, PhD
Order of Authors Secondary Information:	
Abstract:	Social robots have limited social competences. This leads us to view them as depictions of social agents rather than actual social agents. However, people also have limited social competences. We argue that all social interaction involves the depiction of social roles and that they originate in, and are defined by, their function in accounting for failures of social competence.

1: Target Article Authors: Herbert H. Clark and Kerstin Fischer

2. Word Counts:

Abstract: 59

Main text: 987

References: 233

Entire Text: 1508

3. Commentary Title: “Who’s there?”: Depicting Identity in Interaction

4: Full Names:

Professor Patrick G.T. Healey (1).

Dr Christine Howes (2)

Professor Ruth Kempson (5)

Dr Greg Mills (3)

Professor Matthew Purver (1)

Professor Eleni Gregoromichelaki (2, 5)

Dr Arash Eshghi (4)

Dr Julian Hough (1)

5. Institution: Queen Mary University of London.

6. Institutional Affiliations.

(1) Queen Mary, University of London

(2) University of Gothenburg

(3) University of Groningen

(4) Heriot-Watt University

(5) King’s College London

7. Email addresses:

p.healey@qmul.ac.uk.

christine.howes@gu.se

ruth.kempson@kcl.ac.uk

g.j.mills@rug.nl

eleni.gregoromichelaki@gu.se

A.Eshghi@hw.ac.uk

m.purver@qmul.ac.uk

j.hough@qmul.ac.uk

8. URL:

<http://cogsci.eecs.qmul.ac.uk>

<https://gu-clasp.github.io/people/ruth-kempson/>

9. Abstract (59 words):

Social robots have limited social competences. This leads us to view them as depictions of social agents rather than actual social agents. However, people also have limited social competences. We argue that all social interaction involves the depiction of social roles and that they originate in, and are defined by, their function in accounting for failures of social competence.

10. Main Text (987)

Clark and Fischer (C&F) provide a timely reminder that there is a large underappreciated gap between the ambitions of social robotics and the actual social competence of robots (Park et. al. 2021). As they demonstrate, natural conversation presents complex challenges beyond current engineering (e.g. Healey, 2021). Nonetheless, they also point to parallels in the ways in which people interact with each other and with social robots.

This commentary is not concerned with critiquing current social robotics but the ontological distinction underlying C&F's discussion. Specifically, does their account of depiction provide a principled basis for their argument that depictions of social agency fundamentally differ from actual social agency?

C&F discuss varied examples of depictions of social agents including Laurence Olivier's performance of Hamlet. Depiction in these examples is complex. The character -Hamlet- is based on a mixture of characters from earlier plays (possibly also Shakespeare's son); there are multiple versions of the text of Hamlet; different productions select different parts of those texts, different actors perform those parts differently; direction, costume, staging, scenography vary, etc. C&F embrace this complexity, using it to characterise

various aspects of ways people treat interaction with social robots as performance.

The problem, as we see it, is that C&F's account of depiction is so rich, encompassing so much of human social interaction, that the distinction between actual social agents and depictions of social agents dissolves. They show that there are familiar contexts in which people perform a role, such as hotel receptionist, which also involve derived authority, particular communicative styles and particular costumes and props. These roles are depictions and successful interaction in these cases requiring that we recognise and engage with the performance. However, arguably, all human social interaction has these properties. It was Goffman's (1959) insight that this kind of performative, depictive, dramaturgical description can be applied to any human social interaction.

For example, when the receptionist in C&F's example (Section 8.1) switches to being someone who grew-up in the same region as Clark, this is, in Goffman's terms, a switch from one kind of performed identity to another. It involves, for example, switching to certain kinds of community-specific knowledge, norms and patterns of language use (see also Clark, 1996). People have multiple overlapping identities, all involving elements of depiction: different social repertoires, forms of authority, and conventions of interpretation. Moreover, it is unclear why such performances of identity involve depictions rather than *indices* to contextual features that transform the current situation to a new one where terms of the interaction have changed.

Despite this, we share the intuition that the features of interaction that C&F highlight are important. However, the crucial role that they assign to inference and pretence seems overly individualistic, presenting the role of potentially highly sophisticated robots as entirely passive, and ignoring efforts people make to scaffold the interaction. Our suggestion is that one way to retain a meaningful, explanatory role for depictions is to abandon the assumption of any fundamental discontinuity between authentic and performed social agency, instead looking at how depiction functions in interaction. Specifically, we argue that depictions are a means of transforming the relation between interlocutors when social performances threaten to break down; they are invoked to account for the gap between a represented social role and the role invoked to explain the performative failure. Returning to C&F's receptionist example, the inability to provide local hotel information leads to the discovery of the receptionist's actual location which prompts the conversation to switch from 'customer'- 'receptionist' to 'people from Rapid City'.

Not all failures emerge at the level of social performance. When we encounter contemporary social robots, there are a variety of ways in which things can go wrong and a variety of stances we can take to explain the failure (cf. Dennett, 1987). We quickly discover the limitations of robot social affordances and this forces us to reason about, for example, who made this thing? (authority) what

is it supposed to do? (intention/character) is there hardware failure (base scene)? Note that this applies equally to humans and robots: we sometimes invoke problems with authority (e.g. someone is too junior or too young to answer) and hardware problems (someone can't hear, or is too drunk).

There are empirical advantages to approaching depiction in this way. It restricts the range of possible depictions to specific episodes of interaction. Although we can imagine indefinitely many possible social roles, it is only ones that actually figure in interaction that matter. It also provides an index of social competence. The relative frequency with which we invoke interactive depictions or, e.g., hardware problems, provides a measure of how sophisticated a social agent is. Embarrassment accompanies the failure of social roles (Goffman, 1967); involving characteristic displays such as blushing, averting eye contact, face touching and smiling and laughter. Unlike shame, embarrassment also directly implicates other participants in a coordinated understanding of what has failed, how it failed and how to recover from it. Interestingly, robots are not currently designed to systematically recognise or produce signals of embarrassment (Park et. al. 2021).

Our assumption is that what makes an 'authentic' social interaction is the ability to detect and recover from failure - something in principle achievable by machines. Machines can participate in interactions where cognitive abilities are distributed across multiple agents and each can compensate for the failures or inadequacy of the other. The centrality of miscommunication (and ability to recover from it) in human-human interaction (Healey et al., 2018) follows from the observation that we never share the same language, skills or information as anyone we nevertheless successfully interact with (Clark, 1996). This is obvious in, for example, parent-child or expert/non-expert interactions, but actually characteristic of all social exchanges, including interactions with social robots. At present the potential possibilities for divergences may be broader and along different dimensions but this is not, we argue, different in kind.

11. Acknowledgements (optional)

12. Funding statement (see above in Target Article instructions)

Howes was supported by two grants from the Swedish Research council (VR) 2016-0116—Incremental Reasoning in Dialogue (IncReD) and 2014-39 for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

13. Conflicts of Interest statement (see above in Target Article instructions)

None.

14. Alphabetical reference list (APA standard).

Clark, H. H. (1996). Communities, commonalities. Rethinking linguistic relativity, (17), 324.

Dennett, D. C. (1987). The intentional stance. MIT press.

Eshghi, A., Howes, C., Gregoromichelaki, E. (to appear) Action coordination and learning in Dialogue In: Jean-Philippe Bernardy, Rasmus Blanck, Stergios Chatzikyriakidis, Shalom Lappin, and Aleksandre Maskharashvili (eds.) Probabilistic Approaches to Linguistic Theory, CSLI

Goffman, E. (1959). The presentation of self in everyday life London. Allen Lane.

Goffman, E. (1967) Interaction Ritual: Essays on Face-to-Face Behavior (1 ed.). Doubleday, Garden City, NY.

Goodman, N. (1968). Languages of Art: An Approach to a Theory of Symbols. The Bobbs-Merrill Company. Inc. New York, Indianapolis.

Healey P.G.T. (2021) "Human-Like Communication" in Stephen Muggleton and Nick Chater (eds.) *Human-Like Machine Intelligence*. pp.137-151. Oxford University Press.

Healey, P. de Ruitter, J.P, Mills, G.J. 2018 Editors introduction: Miscommunication. *Topics in Cognitive Science* 10(2): 264-278

Howes, C. & Eshghi, A. (2021). Feedback relevance spaces: Interactional constraints on processing contexts in Dynamic Syntax. *Journal of Logic, Language and Information*, 30(2), 331-362

Kempson, R., Cann, R., Gregoromichelaki, E. & Chatzikyriakidis, S. (2016). Language as Mechanisms for Interaction. *Theoretical Linguistics*, 42(3-4), 203-276.

Park, S., Healey, P.G.T. & Kaniadakis, A. (2021) "Should Robots Blush?" In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 717, 1–14. DOI:<https://doi.org/10.1145/3411764.344556>