

# Nonverbal Communication in Human-AI Interaction: Opportunities and Challenges

Joshua Newn, Ronal Singh, Fraser Allison,  
Prashan Madumal, Eduardo Velloso and Frank Vetere  
School of Computing and Information Systems, The University of Melbourne,  
Melbourne, Australia  
newnj@unimelb.edu.au, rr.singh@unimelb.edu.au,  
f.allison@unimelb.edu.au, madumalp@unimelb.edu.au,  
velloso@unimelb.edu.au, f.vetere@unimelb.edu.au

## Abstract

In recent years, we have explored the use of *gaze*—an important nonverbal communication signal and cue in everyday human-human interaction—for use with AI systems. Specifically, our work investigated whether an artificial agent, given the ability to observe human gaze, can make inferences on intentions, and how aspects of these inferences can be communicated to a human collaborator. We leveraged a range of humancomputer interaction techniques to inform the design of a gaze-enabled artificial agent that can predict and communicate predictions. In this paper, we include a snapshot of how AI and HCI can be brought together to inform the design of an explainable interface for an artificial agent. To conclude, we outline the challenges we faced when designing AI systems that incorporate nonverbal communication stemming from our work.

---

### How to cite this book chapter:

Newn, J., Singh, R., Allison, F., Madumal, P., Velloso, E. and Vetere, F. 2020. Nonverbal Communication in Human-AI Interaction: Opportunities and Challenges. In: Loizides, F., Winckler, M., Chatterjee, U., Abdelnour-Nocera, J. and Parmaxi, A. (eds.) *Human Computer Interaction and Emerging Technologies: Adjunct Proceedings from the INTERACT 2019 Workshops*. Pp. 221–226. Cardiff: Cardiff University Press. DOI: <https://doi.org/10.18573/book3.ad>. License: CC-BY 4.0.

## Keywords

Human-AI Interaction · Explainable Interfaces · Nonverbal Communication · Multimodal Input · Intention Recognition · Gaze Input

## 1 Overview

Imagine walking up to a group of peers playing a competitive board game around a table (as shown in Scenario 1 below). You start to observe the situation, the actions and the behaviours of individual players; curious to see which player has the upper hand by inferring their potential plans. The player closest to you asks your opinion as to what might happen in the next rounds. Based on your observations, you will be able to provide inferences to some degree of accuracy and subsequent reasons to why you think they might occur. Now imagine a scenario where humans and AI systems (or robots as depicted in Scenario 2) are playing the same board game together. If the AI systems are able to make the same observations as you (a human spectator) in the previous scenario, would the AI systems make similar inferences? Would these inferences be accurate and timely? Would they be able to explain how they have arrived at their deductions? What information and how much information should such intelligent systems include? Our published and ongoing body of work explores such questions from the perspective of ‘gaze awareness’—if intelligent systems can observe where humans are looking and understand the gaze behaviours within the context, would they be able to improve their interactions with their human counterparts better?

The availability of affordable and improved sensor technologies such as eye-trackers used in our work, combined with our collective experience in designing and conducting HCI and AI studies has presented us opportunities to investigate the incorporation of natural human inputs for Human-AI collaboration. Our initial work focused on understanding gaze in human-human interaction [10, 12], especially for gaze-based intention recognition. We conducted these studies within the context of strategic games and collected rich data using a variety of HCI methods. We found that gaze-based intention recognition is especially beneficial in strategic planning scenarios, allowing players to adapt their strategies preemptively. To elaborate, if a player is able to make accurate and early



inferences on the opponent's plans afforded by observing the opponent's gaze, the player can adjust their own strategy according to the predictions if necessary.

Through the findings and data from our human-centred studies, we developed an artificial agent that combines gaze and planning for human intention recognition [13]. Our gaze-aware agent uses a 'white-box' approach that allows us to understand the underlying algorithms and data structures, which makes it simpler to interrogate the model and its predictions. Our latest paper, forthcoming at INTERACT 2019 [11], evaluates the intention-aware agent in a dynamic collaboration setting. Our findings contribute to the understanding of how researchers can support Human-AI teams through a number of considerations when designing collaborative agents with intention-aware capabilities, including information presentation, context-awareness and explainable agency. The paper highlights the importance of nonverbal communication in Human-AI interaction and provides a general approach for applications where knowing the intentions of others are important for effective interaction (e.g. air traffic control, wargaming). In essence, our research so far serves as the first step towards addressing prerequisites for *man-computer symbiosis* outlined by Licklider in 1960 [7].

## 2 Case Study

As part of our forthcoming paper [11], we designed a study to determine *how* humans formulate predictions and subsequently explain their reasoning process when shown a visual representation of gaze of an opponent in a strategic game.<sup>1</sup> We recruited 20 participants ( $M = 25$ ,  $SD = 3.7$ ) with high proficiency in English.

In this study, we employed an 'inverted' Wizard-of-Oz protocol. In a typical Wizard-of-Oz study, a researcher secretly plays the role of the computer system while a participant interacts with it. In our variation, we asked the participant to play the role of the computer system, and the secret is that there is no enduser. The benefit of this is that it allows us to directly collect a large number of different messages that reflect how the participants think the computer 'should' communicate in an assistive fashion. We posed no restrictions on the language format participants could use for communication, allowing them to freely formulate their messages as long as each message contained a *prediction* of their opponent's intentions followed by an *explanation*. At the end of the study, the participant was given a short questionnaire on their experience, followed by a brief interview based on their responses and communication strategies employed.

We elicited a variety of messages through a well-defined protocol that reinforced the participants' belief in the deception and familiarised them with the task, where they were asked to use a chat application. In our analysis, we found that the ability to successfully formulate messages depended on several factors, including individual ability, experience with the game, the communication strategy adopted and the details of the game recording that was shown.

---

<sup>1</sup> Ticket to Ride—<http://www.daysof wonder.com/tickettoride/en/>.

Participants provided a wide range of explanations for their predictions. We found that complex explanations contain *spatial*, *temporal* and *quantitative* properties, in line with findings using expert explainers [4]. Simplistic explanations, on the other hand, typically described observed behaviours and often only with one property (e.g. “*The opponent was looking at those routes.*”). In order to build a general model, we turn to Malle and Knobe [9]’s explanation model for labelling the properties for more complex explanations elicited with the assumption that the model can be generalised to explain human nonverbal or combined inputs.

Our results show that participants formed explanations from different sources of information available to the agent, such as gaze and actions. Explanations can also include information about past and potential future actions derived following Malle and Knobe [9]’s model. This involves *Causal History of Reasons*, defined as  $O^a$ , and *Intentional Action*, defined as  $I^a$ . Participants showed a strong reliance on gaze to explain their predictions. We believe that gaze being ‘always on’ [6] became more prominent throughout the game for enabling predictions as compared to observable game actions. For this reason, we shall include gaze ( $O^g$ ) as part of every explanation generated using our piece-wise function below.

$$Explanation = \begin{cases} O^g, O^a & \text{if ontic actions observed} \\ O^g, I^a & \text{if intentional action likely} \\ O^g, I^a, O^a & \text{otherwise} \end{cases} \quad (1)$$

Below is an example that combines all three sources of information using our function, forming a prediction with an explanation that is highly detailed:

*“The opponent is building a route from Washington to New Orleans through Nashville in the South East [Prediction (i)]. The opponent has claimed part of this route [ $O^a$ ], has been looking at the routes between Raleigh and Little Rock repeatedly [ $O^g$ ] and is likely to claim Nashville to Raleigh next [ $I^a$ ].”*

In summary, this study presents a simple case of how human-centred approaches from HCI can be used to inform the design of explainable interfaces. The results from this study form the basis for a computational model of explanation, in which we can use gaze and ontic actions to form explanations, and we can vary the level of detail as needed. Beyond answering the *how* and *what* questions to meet our design goals, we learn it is crucial to know *when* (or how often) to provide an explanation in the context of predictions, and this requires the agent to be contextually-aware of the what the assisted-player already knows and whether the information to be communicated is helpful to them. Lastly, we also learn that it is possible and essential to consider the portrayal of uncertainty when communicating predictions as used in natural language, providing an alternative to using confidence levels as used in traditional AI systems.

### 3 Opportunities & Challenges

The case study presented in this paper is just one example of how we have utilised a human-centred approach to inform the design of AI systems, which has subsequently led us to better augment the agent's ability to detect human intention from gaze. Hence, we posit that for AI to work with their human collaborators effectively, AI systems first need to harness nonverbal cues commonly present in human-human interaction. Since, we have expanded our work to explore other nonverbal inputs (e.g. gestures, facial expression) for Multimodal Human-Agent Collaboration.<sup>2</sup> Simultaneously, we have continued to use the combination of AI and HCI in our work, such as to develop and further evaluate a general dialogue model for explanations by putting AI-assisted humans in the loop [8]. At present, our work focuses on the adoption of nonverbal communication in Human-AI interaction and is situated at the crossroads of addressing the design aspects (e.g. [2, 5]), overcoming the technical challenges (e.g. [3]), and the existing work on nonverbal communication in human-robot interaction (e.g. [1]).

However, many challenges remain until we can understand how to utilise nonverbal inputs fully. In the first place, it is often difficult to find a suitable use case to investigate that fully demonstrates benefits from Human-AI integration. In our work we were challenged to think differently due to the nature of gaze as a subtle and often unnoticed signal; it required the use of HCI to build an understanding of how humans utilise gaze before we could design a system that performs similarly or better. In the context of building explainable AI interfaces, we aim to tackle some immediate challenges, such as by determining the proper explanation interface and medium (e.g. visual, verbal, textual explanations). Perhaps the most prominent challenge faced is to ensure that the models that integrate multimodal input can be generalised to other contexts. Nevertheless, our work presents the first step towards our goal of building explainable agents that can assist, mediate or negotiate with knowledge of multiple users' intentions.

### References

1. Admoni, H., Scassellati, B.: Robot nonverbal communication as an ai problem (and solution) (2015)
2. Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P.N., Inkpen, K., Teevan, J., Kikin-Gil, R., Horvitz, E.: Guidelines for human-ai interaction. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 3:1–3:13. CHI '19, ACM, New York, NY, USA (2019). <https://doi.org/10.1145/3290605.3300233>
3. Bednarik, R., Eivazi, S., Vrzakova, H.: A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data, pp. 111–134. Springer London, London (2013)

---

<sup>2</sup> <https://cis.unimelb.edu.au/agentlab/human-agent-collaboration/>.

4. Dodge, J., Penney, S., Hilderbrand, C., Anderson, A., Burnett, M.: How the experts do it: Assessing and explaining agent behaviors in real-time strategy games. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 562:1–562:12. CHI '18, ACM, New York, NY, USA (2018)
5. Eiband, M., Schneider, H., Bilandzic, M., Fazekas-Con, J., Haug, M., Hussmann, H.: Bringing transparency design into practice. In: 23rd International Conference on Intelligent User Interfaces. pp. 211–223. IUI '18, ACM (2018)
6. Jacob, R.J.K.: What you look at is what you get: Eye movement-based interaction techniques. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 11–18. CHI '90, ACM, New York, NY, USA (1990)
7. Licklider, J.C.R.: Man-computer symbiosis. IRE Transactions on Human Factors in Electronics **HFE-1**(1), 4–11 (March 1960)
8. Madumal, P., Miller, T., Sonenberg, L., Vetere, F.: A grounded interaction protocol for explainable artificial intelligence. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 1033–1041. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019), <http://dl.acm.org/citation.cfm?id=3306127.3331801>
9. Malle, B.F., Knobe, J.: Which behaviors do people explain? a basic actor–observer asymmetry. *Journal of Personality and Social Psychology* **72**(2), 288 (1997)
10. Newn, J., Allison, F., Velloso, E., Vetere, F.: Looks can be deceiving: Using gaze visualisation to predict and mislead opponents in strategic gameplay. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 261:1–261:12. CHI '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3173574.3173835>
11. Newn, J., Singh, R., Madumal, P., Velloso, E., Vetere, F.: Designing interactions with intention-aware gaze-enabled artificial agents. In: Human-Computer Interaction – INTERACT 2019. pp. 255–281. Springer International Publishing, Cham (2019). [https://doi.org/10.1007/978-3-030-29384-0\\_17](https://doi.org/10.1007/978-3-030-29384-0_17)
12. Newn, J., Velloso, E., Allison, F., Abdelrahman, Y., Vetere, F.: Evaluating realtime gaze representations to infer intentions in competitive turn-based strategy games. In: Proceedings of the Annual Symposium on Computer-Human Interaction in Play. pp. 541–552. CHI PLAY '17, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3116595.3116624>
13. Singh, R., Miller, T., Newn, J., Sonenberg, L., Velloso, E., Vetere, F.: Combining planning with gaze for online human intention recognition. In: Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems. pp. 488–496. AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2018), <http://dl.acm.org/citation.cfm?id=3237383.3237457>