Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents

Joshua Newn, Ronal Singh, Fraser Allison, Prashan Madumal, Eduardo Velloso, Frank Vetere

Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents

Joshua Newn[0000−0001−5769−6297], Ronal Singh[0000−0003−3352−0486],
Fraser Allison[0000−0002−7005−7390], Prashan Madumal[0000−0002−2466−670X],
Eduardo Velloso[0000−0003−4414−2249], and Frank Vetere[0000−0003−3727−9204]

School of Computing and Information Systems,
The University of Melbourne, Australia
{newnj,rr.singh,f.allison,madumalp,evelloso,f.vetere}@unimelb.edu.au

Abstract. As it becomes more common for humans to work alongside artificial agents on everyday tasks, it is increasingly important to design artificial agents that can understand and interact with their human counterparts naturally. We posit that an effective way to do this is to harness nonverbal cues used in human-human interaction. We, therefore, leverage knowledge from existing work on gaze-based intention recognition, where the awareness of gaze can provide insights into the future actions of an observed human subject. In this paper, we design and evaluate the use of a proactive intention-aware gaze-enabled artificial agent that assists a human player engaged in an online strategy game. The agent assists by recognising and communicating the intentions of a human opponent in real-time, potentially improving situation awareness. Our first study identifies the language requirements for the artificial agent to communicate the opponent's intentions to the assisted player, using an inverted Wizard of Oz method approach. Our second study compares the experience of playing an online strategy game with and without the assistance of the agent. Specifically, we conducted a within-subjects study with 30 participants to compare their experience of playing with (1) detailed AI predictions, (2) abstract AI predictions, and (3) no AI predictions but with a live visualisation of their opponent's gaze. Our results show that the agent can facilitate awareness of another user's intentions without adding visual distraction to the interface; however, the cognitive workload was similar across all three conditions, suggesting that the manner in which the agent communicates its predictions requires further exploration. Overall, our work contributes to the understanding of how to support human-agent teams in a dynamic collaboration scenario. We provide a positive account of humans interacting with an intention-aware artificial agent afforded by gaze input, which presents immediate opportunities for improving interactions between humans and agents and demonstrates the usefulness of human elicitation for producing communication protocols for effective interaction.

Keywords: Human-AI Interaction·Intention Recognition·Explainable Interface·Human-AI Teaming·Intention Awareness·Eye Tracking·Gaze Awareness.

1   Introduction

Gaze is an important nonverbal communication signal in everyday human-human interaction [4], and has become a popular research topic for technology-mediated interaction [17,

43, 59]. The ability to tell what someone is looking at—'gaze awareness'—is a useful way to gauge the attention of others [1, 2, 14, 62]. Gaze observed over time has been shown to be an effective predictor of human intention [26, 27, 49, 55]. A common approach for gaze awareness is to visually overlay a user's gaze over a shared interface, which provides others rich insights into the mind of the tracked user. This complementary layer of communication has numerous benefits such as improved coordination [2, 12, 14] and situation awareness [49]. Despite these benefits, overlaying gaze on the interface can add a highly distracting element to the task at hand [49], confuse users when there is a mismatch with other modes of communication such as speech [14], and scales poorly with multiple users.

In this paper, we explore how an artificial agent that interprets eye movements can alleviate issues commonly associated with visual gaze awareness, and how humans respond to agent-derived intentions from gaze and observable actions. A socially interactive agent that can understand human gaze can potentially improve the interaction with its human counterparts [27, 55], such as by adapting its behaviour to their anticipated intentions, or even support the user by communicating the intentions of others. However, much investigation is still needed from an interaction design perspective before humans can work alongside such agents effectively, with each counterpart playing to its strengths.

Our work presents a step towards artificial agents that can interpret and communicate human intentions based on nonverbal behavioural signals. We designed and evaluated the communication protocols of a proactive gaze-enabled artificial agent for communicating intentions to a human player in the context of an online strategy game. The agent assists by making inferences about the opponent's intentions based on their gaze and actions, allowing the user to focus on formulating better strategies with improved awareness of the situation. By abstracting the gaze data into a written prediction of what the opponent intends to do, we avoid the distracting nature of gaze visualisation as found in past research [14, 49, 50, 62]. As nonverbal cues are difficult to articulate, our first step was to build a linguistic model of intention recognition derived from human observers. This process resulted in a general model of intention communication, which we incorporated into the artificial agent.

Our following step evaluates the model while comparing the existing approach of using gaze visualisation to infer intentions in strategic gameplay to our proposed approach of abstracting gaze input into a written prediction through an artificial agent. We designed a within-subjects study with three conditions with varying levels of information provided to the player. In the first condition, we provide players with a real-time visualisation of their opponent's gaze, allowing them to interpret the information as they see fit. In the second condition, the agent sends the player inferences about the opponent's plans, followed by an explanation of the observed behaviours that it used to form the prediction in an attempt to be transparent about its reasoning process. In the third condition, the agent sends the player predictions about their opponent's plans without an explanation of its reasoning process, to allow the player to have their own beliefs about the agent's logic, and without any direct knowledge of the data that led to that inference.

We conducted the within-subjects study with 30 player pairs, in which the evaluated players reported a positive experience when engaging with the agent in terms of preference and usefulness for situation awareness, and a perceived reduction in cognitive workload and distraction. Though our results show that the agent can facilitate awareness of another user's intentions without adding visual distraction to the interface, there was no significant difference in the players' cognitive workload for the written

prediction conditions as compared to the live gaze visualisation condition. This suggests that the manner in which the agent communicates requires further exploration.

All in all, our work presents two primary contributions towards the design of artificial agents that collaborate with humans, from both sides of the interaction. First, we demonstrate that proactive artificial agents can support their human counterparts by observing gaze to provide support through intention recognition and that it is possible to communicate these predictions effectively in written language. Second, our work shows that the human counterpart can be supported by a proactive agent that communicates intentions, which maintains situation awareness while reducing visual distraction when compared to using a live gaze visualisation approach

2 Related Work

2.1 Shared Gaze Awareness

Gaze visualisation is by far the most common approach for utilising gaze input in technology-mediated human-human interaction. This approach provides a complementary layer of nonverbal communication, especially beneficial in remote settings where users cannot see where other the users they are interacting with are looking. Observers can derive rich information from gaze behaviours displayed over an interface (e.g. scanning, focus on an object, and repeated comparisons of different objects [56]). These gaze behaviours provide clues about the other person's cognitive processes, i.e. the ability to discern their intentions [50]. The benefits are well demonstrated in multiuser scenarios, improving communication and coordination in collaborative settings (e.g. [2, 12, 24, 62]). Gaze visualisation has also been explored in competitive gameplay [58], highlighting its potential for increasing social presence between remote players [36, 45], and for enabling players to recognise the intentions of others in real-time [49, 50].

Despite its numerous benefits, researchers have commonly found that using live gaze visualisation can be 'distracting' and 'confusing' for an observer to interpret [14, 50, 62]. We believe this is because humans are not accustomed to interpreting visual representations of gaze, as the focal point of gaze is 'invisible' in normal everyday interpersonal interaction [49], and that an added layer of continuous information draws the user's attention away from the task at hand when displayed. As gaze visualisation is highly dependent on context and individual preference [15, 49], software for visualising gaze in real-time often allows users to control its parameters such as by adjusting the colour, opacity and smoothness [7, 13]. The recent release of Tobii Ghost[1]—a free commercial software designed to allow eSports audiences to view customised gaze visualisations of players in real-time—further exemplifies the growing popularity of this feature in gaze visualisation applications.

2.2 Gaze-Based Intention Recognition

Though human attention can be easily inferred by the direction of a person's gaze, discerning their intention through their gaze is a far more complicated process. The observer must distinguish between intentional and unintentional behaviours, and gaze direction alone provides very few clues to do so. In our previous study, we demonstrated that using an aggregated visualisation of gaze can enable human-human intention recognition in competitive gameplay, with benefits such as early inference of intentions

---

[1] https://tobiigaming.com/software/ghost/

[49]. Despite such benefits, the study found that players who could see the gaze of their opponent had no gain in performance, due to its cost in time and attention—by attempting to infer their opponents' strategies, they ended up neglecting their own. Players who did manage to reach a balance stated that the broad clues provided by gaze awareness were beneficial to formulating and adjusting their strategy. For instance, they could ignore certain areas of the game-board if they noticed that their opponent had not looked there. Overall, these findings suggest that effectively managing the cognitive demands of inferring the opponent's strategy and devising one's own is the key to successfully making use of the opponent's gaze information.

However, it is unlikely that humans can fully operationalise gaze while performing complex tasks without assistance, due to the limits of human working memory. As visual behaviour is intrinsically linked to how humans plan and execute actions [34], researchers have explored the use of computational techniques to perform intention recognition from gaze, typically employing a machine learning approach (e.g. [5, 26]). In a previous paper, we proposed an alternate approach that incorporates visual behaviour into model-based intention recognition using automated planning [55]. We leveraged the fact that humans plan ahead in strategic scenarios and that the incorporation of gaze as priors in a planning-based model resulted in the computation of predictions with high accuracy, earlier and with no computational cost when compared with a base planning model that did not use gaze input. Overall, such works, exemplify the use of computational techniques to harness the rich information available from the observation of gaze behaviour.

2.3   Human-Agent Teaming

In 1960, Licklider proposed the vision of *man-computer symbiosis*, where computers would be able to work with humans to solve problems that are not easily addressed if attempted by either counterpart individually [38]. For instance, while computers can perform complex calculations and repetitive tasks far better than humans, humans are better at visual-spatial reasoning and at exercising judgement. However, enabling this symbiosis through mixed teams comes with significant challenges to effectiveness [33]. One such challenge is a lack of agent transparency, which hinders the human partner's ability to understand the decisions of the artificial agent [30, 47]. The lack of transparency can lead to adverse effects for the human partner, such as a reduction in trust when working together, and therefore, a potential for disuse [6, 10, 35, 61].

Researchers in AI argue that providing explanations supports transparency and may improve trust in the system [23, 41, 46, 51, 61]. Moreover, when using an agent as a decision aid, users would often seek an explanation of its output to improve their own decision making [60]. However, for an explanation to be effective, it must be at the right level of detail [31]. An explanation of how something works will fail if it presupposes too much and skips over essential information, or if it provides a level of detail that leads to an increase in cognitive workload and an associated decrease in effectiveness [51]. Further, we need to consider the application domain, the audience of the explanation [21], as well as the *presentation format* (how to explain) and the *content* (what to explain) [19, 31].

From a different perspective, dissimilarities between human language and computer language pose another consideration for real-time cooperation, which Licklider states "*may be the most serious obstacle for true symbiosis*" [38]. The author explains that humans think more naturally and easily in terms of goals than specific itineraries, implying the existence of human goals during communication. Computers, however, communicate better in terms of procedural instruction (i.e. courses), which may be

redundant or not meaningful to a human collaborator.

In summary, there are numerous benefits for implementing gaze input for computer-mediated interaction afforded by advances and availability of eye-tracking technology. However, information overload, interpretation difficulty and scalability using the conventional approach of gaze visualisation hinder its full potential for multiuser settings. Recent work in AI has shown intelligent agents have the potential to perform intention recognition from gaze input, which is often a complex task in human-human interaction, especially when the user is already preoccupied. Our work intersects these areas by using a collaborative intelligent agent to support a human counterpart by recognising the intentions of others based on their gaze for them. To do this effectively, we must first consider how an ideal agent would communicate intentions once recognised, addressing the language mismatch prerequisite. Second, we must consider an agent's explanation capabilities to support transparency, where the agent can provide insights into its reasoning process to gain the user's trust. Lastly, we need to consider the optimum level of support as different levels of AI support can result in changes in cognitive workload, positively or negatively [9].

3   Research Design

From our review of the literature, an ideal intention-aware agent for human-agent interaction in the context of teaming should possess the following capabilities:
1. Infer a users' intentions accurately based on gaze observation and other available sources of information (e.g. observable explicit actions) in a timely manner.
2. Communicate inferred users' intentions to an assisted user in a way that the user finds easy to understand, such as through natural language.
3. Increase the users' situation awareness while reducing the users' cognitive workload (in comparison with current approaches such as the use of gaze visualisation).

We conducted two studies to evaluate the prospects of an agent possessing these capabilities. Our first study identified the language that humans use to describe the intentions of third parties over short text-based messages. The findings from the study provided the language requirements for our artificial agent. In the subsequent study, we evaluated our enhanced artificial agent with participants using an online strategy game. We obtained ethics approval from our University's ethics committee, as both studies involved mild deception of the participants. Both studies required a scenario in which participants were required to deduce another person's intentions through a computer system. For our purposes, we used the digital version of an online competitive turn-based strategy board game called *Ticket to Ride*. In this game, players compete to build connections between cities based on drawn 'ticket' cards (e.g Dallas to New York). The core of the game is to keep their intentions hidden as an opposing player can gain a significant advantage by correctly guessing their hidden plans. Therefore, players must plan their routes carefully to minimise the risk that an opponent will guess their intentions and block them by claiming the routes that they need first. More detailed information on the rules of *Ticket to Ride* can be found on the game's website[2].

4   Study 1: Language Identification

In this study, we developed an effective language model for a gaze-aware artificial agent to communicate an opponent's intentions to a user through text, and conducted a

---
[2] http://www.daysofwonder.com/tickettoride/en/usa/

study to generate specific language data for our broader scenario. We used a variation of the Wizard of Oz prototyping method in which participants played the role of the 'artificial agent' to produce language according to what they think is appropriate to the task, instead of the language being determined by the researchers or system designers.

The goal of our artificial agent is to promote the user's *situation awareness*, defined by Endsley [20] as: "*perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the future*". Gaze awareness has been shown to be especially beneficial for situation awareness, particularly when a player is in a strategic game can make correct inferences about their opponent's strategy early in the game [49]. However, prior literature on agent transparency in general tasks indicates two important aspects of agent communication: *presentation format* and *content* [19, 31]. We used the Situation Awareness-Based Transparency Model (or SAT Model) [11] as a model of agent transparency to support a user's situation awareness. In this model, the agent communicates different types of information at three levels to support the user. At the lowest level, the agent communicates its own state, which includes the agent's intentions. At the middle level, the agent communicates information regarding its reasoning process, and at the top level, the agent communicates information regarding potential future states.

For this study, we recruited 20 participants (11M/9F) from The University of Melbourne, aged between 20 and 32 years (M=25, SD=3.7), to take on the role of a 'predictor-explainer'. We selected participants based on their self-rated English proficiency in our recruitment questionnaire, as we required participants to produce a rich vocabulary around gaze behaviours, observable actions and the communication of intentions. We provided participants with the rules of the game at the time of recruitment, and we compensated them with a $15 (AUD) gift card upon completion of the study.

4.1   Experimental Setup & Procedure

Upon arrival, we sat the participant in front of a computer and obtained the participants written consent to participate in the study. The participant and experimenter sat at opposite ends of the table so that the experimenter's display was not visible to the participant. Figure 1-Right shows the technical setup consisting of a laptop connected to two 23-inch monitors on a rectangular table, located in a study room. The experimenter then introduced the task by explaining that there were two other players in separate rooms preparing to play *Ticket to Ride* against each other. The experimenter told the participant that they had been randomly selected to take on the role of a 'predictor-explainer' (or appraiser), who would watch the game between the two other players via the computer and send assistive messages to one of them, their 'teammate'.

In reality, there was only one participant in each session (i.e. themselves). The game of *Ticket to Ride* shown to the participant was pre-recorded (actual game), and we used each recording only once. To clarify, we showed 20 different games played by 20 different player pairs from our previous study data set [49]. The recorded player was naive to the fact their gaze was being observed, meaning that the participant observes natural gaze behaviours. We did this for two reasons, (1) in order to elicit a wide range of textual representations from different game scenarios and (2) there was no need for anyone to receive the participant's messages, as the lexical content of those messages was the focus.

The recording of the game included a 'live' dynamic heatmap visualisation of the gaze of the 'opponent' player (as shown in Figure 1). We designed and employed a protocol to continually reinforce the participants' belief that they were engaged in a live online game
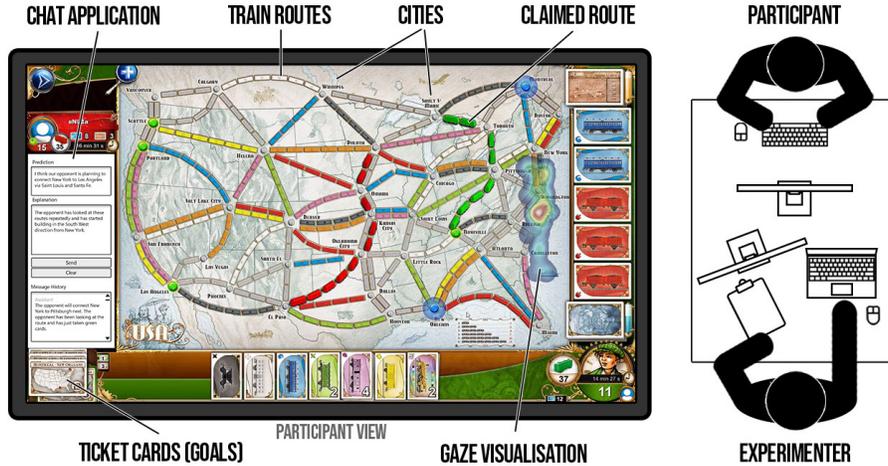
Fig. 1: Left: Participant View. Right: Experimental Setup.

with two other players throughout the study. For example, as each session was designed to last a maximum of an hour, we informed the participant in advance that the game would begin at a fixed time, partway through the session, as all "three" participants needed time to be introduced to the study and familiarise themselves with *Ticket to Ride* through the game's tutorial. The researcher was only allowed to clarify the rules about the game when prompted during the study to avoid any influence on the data.

We describe this approach as 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 [32, 53]. In our study, the participant is asked to play the role of the computer system, and the secret is that there is no end user. 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. A similar approach has been used in the context of machine learning to 'bootstrap' a Reinforcement-Learning-based dialogue system on human-generated activity [54].

Before it was 'time to join the game', the experimenter showed the participant four short clips (introduced as pre-recordings rather than a live game), representing four scenarios with the live dynamic gaze visualisation. This was to start the participant thinking about how they could form *predictions* from the information available, particularly the gaze visualisation, and then from *explanations* in text about their reasoning process. This allowed them to develop confidence in their ability to observe and communicate simultaneously during the live game. We reminded participants to provide messages that their 'teammate' would find helpful, and to build the teammate's trust by being transparent in how they derived their predictions through their explanations.

Next, we demonstrated a simple chat application that served as the means of communication with their teammate (see Figure 1). The application contained two text fields to input their *prediction* and *explanation* respectively, a send button and a window showing the conversation. The application logged all messages sent and included a validation to ensure both text fields are not empty. We augmented the application to select a response from a range of automated natural language responses in reply to every message sent by the participant to keep up the deception. The responses

mimicked a 'busy player': one that replied with a short delay, sometimes did not reply at all, and typically replied with a brief response. The majority of responses consisted of acknowledgements, while the remaining introduced expressions of uncertainty about the participant's messages to convey human-like qualities (e.g. "I don't think so.").

At the prescribed start time, the experimenter streamed the recorded game as if it was a live game feed and informed the participant that the game had started. We posed no restrictions on the syntax, or semantics participants could use for their messages, allowing them to freely formulate them as they saw fit as long as each contained a prediction of their opponent's intentions followed by an explanation for their prediction. At the end of the study, the researchers conducted a short interview with the participant to find their experience embodying the role. Lastly, we debriefed participants about the deception and provided participants with the opportunity to inquire about our objectives.

4.2   Findings

We elicited a total of 249 raw messages (M=12.4 messages per participant), with a high deviation between participants (min=4, max=23). The ability to successfully formulate messages depended on several factors, including individual ability, experience with the game, the communication strategy adopted and the recorded game shown. We discarded messages where participants attempted to communicate with their teammate casually or provided a recommendation instead. However, we included recommendations that resemble a prediction that included a clear explanation (e.g. *"You should block Helena to Duluth, our opponent is likely to claim this the route next as he has repeatedly been looking at it."*). We also split messages that contain two mutually exclusive predictions (e.g. *"The opponent is interested in the west coast. Opponent may build routes around New York."*), which typically occurs when participants formed another prediction while forming an initial unrelated prediction but have the same reasoning process for both. Finally, we obtained a total of 246 messages after our filtering process for analysis.

Prediction Format For the prediction part of each message, we stripped them into its essential meaningful components to obtain a minimal format for predictions (e.g. From [City] to [City] through [City].), which gave us a total of 45 initial formats. We merged formats that were similar in nature into key prediction formats (examples shown in Figure 2), each demonstrating unique characteristics in terms of abstraction. We also noted that participants conveyed their level of confidence when providing their predictions, using words that express uncertainty (e.g. i think/maybe/will try). As studies on explanations argue for showing system uncertainty [3, 39], we will introduce uncertainty when communicating predictions, including stating alternate routes when the likelihood of the plan is similar (e.g. To [City] or [City] through [City] from [City] or [City].).

Explanation Content 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 [16]. 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 [44]'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. Following the model, explanations can include information about past and potential future actions, i.e. *Causal History of Reasons*, defined as O_A, and *Intentional Action*, defined as I_A. As our logs

showed that participants had a strong reliance on gaze to explain the predictions, we include gaze (O_g) as part of every explanation generated using our piece-wise function below. We believe that gaze being 'always on' [28], becomes a valuable source of information for participants throughout the game, especially when only a few observable actions have been performed by the opponent.

$$Explanation = \begin{cases} O_g, O_A & \text{if ontic actions observed} \\ O_g, I_A & \text{if intentional action likely} \\ O_g, I_A, O_A & \text{otherwise} \end{cases} \quad (1)$$

Therefore, combination of all three sources of information forms an explanation that is *detailed*, for example:

"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]."

Reasoning & Communication Strategies Participants adopted two general strategies for reasoning about and communicating the intentions of their opponent, which they maintained either strategy throughout, or interchanged between the two depending on the situation. We found that the strategies were reflective of the two systems of Kahneman's Dual Process Theory [29]—*System 1* (heuristic, intuitive) and *System 2* (systematic, analytical). The first strategy was to send as many messages as possible, in fear of missing out on communicating predictions that may be important to their perceived teammate. Due to this time pressure, we believe participants adopted System 1, where they made use of their intuition, and that their rate of communication was limited to their typing speed. In contrast, the second strategy was closer towards System 2, where participants took a conscious effort to reason about the opponent's intentions and overall strategy, as they wanted to provide the best possible prediction accompanied by a detailed explanation of their reasoning process. This strategy resulted in fewer predictions, especially if the current prediction or reasoning did not change.

Participants on average generated more predictions at the beginning of the game and followed by fewer predictions towards the end of the game, representing its relevance. Unless the opponent's plan changes, more recent predictions will be less relevant, especially if the new predictions were part of the plan that has already been predicted. In our interviews, participants noted the most difficult aspects of explaining is to come up with the best possible explanation, and also what to communicate when unsure how they have come about the prediction. This is when System 1 (or intuition simply) often comes into play, which makes it hard to quantify certain aspects such as how much the opponent has looked at one part as compared with another. Participants also noted that timely predictions would be most helpful, but this is difficult to tell how far in advance the opponent will perform the predicted action (e.g. in how many turns).

5   Study 2: Evaluation

By combining the language model derived from Study 1 with an instance of an intention-aware artificial agent from our previous work, we can now evaluate the experience of playing an online strategy with and without agent assistance. Figure 2 summarises our experimental setup of two observation rooms and a control room. Each session

involved three researchers, two to facilitate the players and the third, an unseen human to assemble the predictions from the artificial agent into natural language following a set of rules. Both setups were identical for both players, except for the eye trackers where the evaluated player (P_A) was equipped with Tobii Pro X2-30 (for pupillary data) and 'naive' opponent (P_B) was equipped with a Tobii 4C eye tracker.
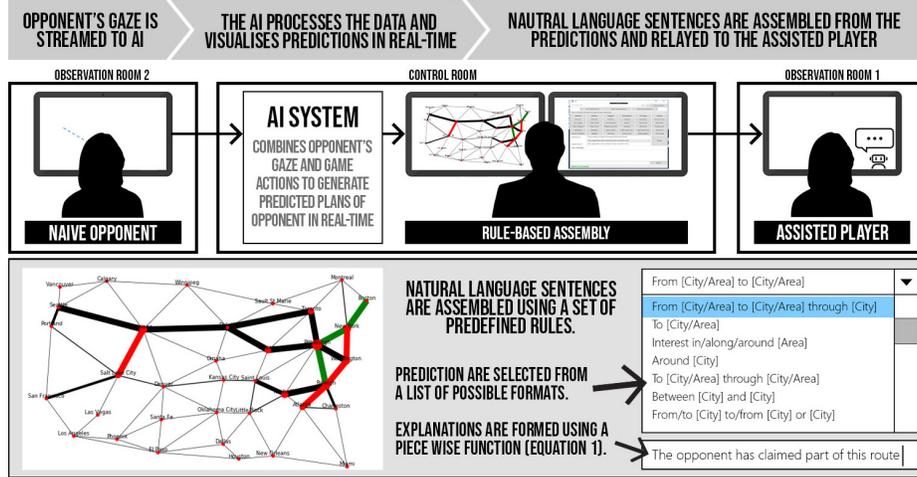

Fig. 2: **Top**: Experimental Setup and Communication Flow. **Bottom**: AI System Visualisation and Assembly Process. Opponents intentions are displayed by increasing the line thickness of routes. The thicker the line, the more likely the route will be claimed. Coloured lines represent the claimed routes (player: green, opponent: red). The size of the city indicates where the opponent has fixated upon (the larger the city, the more the opponent has looked upon).

We recruited 60 players (34M/26F) for the study and allocated them randomly into two equal groups according to gender (17M/13F in each). At the time of recruitment, we informed players that the purpose of the study was to collect physiological data while they played a strategic game. The first group (Group A) consisted of 'aware' assisted players, aged between 18 and 50 years (M=26.9, SD=6.9), while the second group (Group B) consisted of 'naive' players, aged between 18 and 33 (M=25.6, SD=3.8) to be the opponents. 17 assisted players and 10 naive opponents have played the game before. All players were compensated with a \$20 (AUD) gift card for their participation.

5.1   Intention-Aware Gaze-Enabled Artificial Agent

We instantiated an artificial agent that performs intention recognition using the combination of ontic actions and gaze using a planning-based model from our previous work on the subject [55]. The approach 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, and therefore generate explanations when compared to other approaches. Further, research has shown that humans prefer working with an agent through planning; reporting the perceived reduction of cognitive workload, and the ability to maintain situation awareness for short-term tasks [48]. The objective of the agent assistance in this work is not to solve the *Ticket to Ride* game by providing step-by-step recommendations to the assisted player but to explore how an agent can assist a human player by maintaining and communicating its beliefs of an opponent's intentions.

Our decision to adopt an artificial agent instead of a Wizard of Oz approach as used in Study 1 was for three reasons. First, using the data set our previous study [49], the agent scored positively higher compared to a human interpreter using gaze visualisation ($F_1-Score$: 0.57 versus 0.37 respectively) in terms of plan recognition. This means that an assisted player playing alongside the agent would receive more accurate predictions than with a human assistant (or wizard), which gave us confidence in its adoption. Second, the more accurate agent provides better ground truths overall, meaning that even if we provide the opponents goals (destination cities) to the wizard, the system remains far better at discriminating and predicting the most likely plans and can provide this information earlier. This ensured relative consistency of predictions and provides a realistic impression of what such systems can do across participants. Third, as previously found [49], human interpreters can be subject to biases, especially when the human interpreter fixates on incorrect predictions and overlooks other predictions.

As part of this work, we developed a graph visualisation to display the predictions made by the agent to assist in the rule-based assembly stage (shown in Figure 2-Bottom). The graph displays the combination of the top 10 most likely plans of the opponent. The thickness of the edges (representing routes) increases according to the number of times it appears in the top 10 plans; indicating the likelihood of the player choosing that particular plan. Further, the graph not only shows the opponent's plans at a macro-level but also the possible combinations that the opponent may use to achieve their intentions.

5.2   Study Conditions

We designed three conditions representing three levels of information abstraction. At the lowest level (GAZE VIZ), we show the assisted player (P_A) the gaze of the native opponent (P_B) throughout the game using a real-time heatmap (visualisation as shown in Figure 1). This condition allowed players to make their inferences at the cost of their attention and serves as a baseline condition as we displayed the visualisation throughout the game.
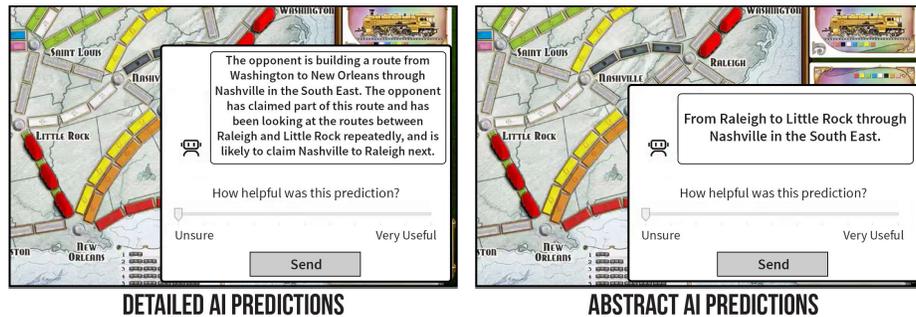

Fig. 3: AI Prediction Examples.

At the mid-level (DETAILED AI PREDS), we assemble the intentions and observed behaviours into our text-based language model informed by Study 1. Here, we presented the prediction as an *Intentional Action* [44]—what the opponent intends to do next while being transparent about its reasoning process. As part of natural language, we conveyed uncertainty when communicating the predictions and provided temporal, spatial and quantitative elements where possible. At the highest level (ABSTRACT AI PREDS), the agent provided an abstract about the predicted plan through one of the formats from our language model. As both AI prediction levels are reflective of the Dual Process Theory [29]

systems and strategies described in Section 4.2, we simulate the communication frequency accordingly. For DETAILED AI PREDS, we require the formation of detailed messages and therefore set the frequency to every 2 minutes so that the system can make a sufficient observation to form the best possible prediction and explanation. For ABSTRACT AI PREDS, the frequency was set to 1 minute, as we only need to send the best possible prediction at that point in time. We counterbalanced the study conditions using a Latin square to minimise any learning effects. As this is a within-subject study design, we only subjected them to the assisted player (P_A), in which we presented as a 'mode of assistance'. For both conditions, the researcher made it explicit to the assisted player that the AI uses their opponent's gaze behaviour and observable game actions to generate predictions.

5.3   Measures & Analysis

To evaluate the player experience in each condition, we designed a repeated-measures questionnaire. As there was no specifically designed questionnaire to measure the experience of performing intention awareness, we formed our questions based on our previous work [49], which measures the subjective experience of players when performing intention recognition with and without gaze visualisation. For each measure, we employed a 7-point Likert scale (1 being full disagreement, 7 being full agreement) and questions include the participant's perceived ability to discern intentions and formulate strategy, the effects of information presented during gameplay (such as whether it has influenced the outcome or have caused them to play differently), and whether the condition presented were distracting and were informative. For the AI conditions, we included two other measures, which asked the players how well they *understood* the AI predictions and how *reliable* the AI performed in predicting the opponent's intentions, and only in the DETAILED AI PREDS condition, we asked players about the *clarity* of the explanations to validate messages formed using our model. At the end of the study, we measured the overall experience of using all three conditions, we asked players to rate the conditions with regards to *preference*, *demand* and *usefulness* from most to least. We then prompted the players on the ratings for each measure as part of our subsequent semi-structured interview.

   To measure cognitive workload unobtrusively, we use the recently proposed Index of Pupillary Activity (IPA) metric [18], which measures the frequency of pupil diameter oscillation. The metric shows a direct correlation with working memory, making it a plausible way to measure cognitive workload. Further, we employed traditional measures of the cognitive workload from eye movement behaviour from prior work (e.g. [9]), such as *long fixations* (i.e. fixations >500 ms), which indicate deeper cognitive processing. We also used NASA-TLX [22] to capture perceived workload based on six subscales—mental demand, physical demand, temporal demand, performance, effort and frustration.

5.4   Participants & Procedures

To manage the complexities of the study, researchers followed a strict rehearsed protocol. Both players were given an initial briefing together upon arrival that explains that we will track their physiological signals throughout the study for post-study analysis. We then provide players with the written overview of the study, consent form and basic demographic questionnaire to fill out before separating randomly into one of the two observation rooms with the allocated facilitator. We instructed the players to play the game's interactive tutorial for up to 10 minutes to get familiar with the game and its controls, regardless of experience. Players then played three rounds of *Ticket to Ride* against each other, with each testing a different study condition. At the start of

each round, we requested each player to pick all three randomly assigned 'ticket' cards for them to attempt to complete (each representing a pair of 'goal cities', potentially having up to six initial goal cities). Players were asked to 'think aloud' during the game about their strategy; their opponent's strategy; what they were thinking and what their opponent might be thinking. Each player was given a 12-minute cumulative time allowance for their total turns in each round to ensure timely completion. If either player ran out of time, we manually calculated the scores for that round. We video recorded the screen and rooms for both players for the entire duration of the session. Each session lasted approximately 120 minutes in total. For the remainder of the section, we describe the procedure for each player separately for clarity.

Player A (Assisted Player) Procedure Once the players entered their respective rooms, the facilitator (F_A) informed the player that they had been randomly selected to be the 'aware' player while making it clear at no point during the study that their information will be exposed to their opponent (P_B). The facilitator then calibrated the player's eyes with the eye tracker using the default calibration before starting the tutorial. We then informed the player that they would play three rounds of the game against player P_B and will receive 'additional information' about their opponent's intentions without their knowledge and will vary according to the condition. In all conditions, the player received prompts with a slider (see Figure 3). The primary purpose of the rating scale is for players to reflect on the information that is being presented to them. The players were instructed to verbalise why they had given a particular rating. At the end of each condition, we administered the NASA-TLX immediately before asking them to fill up a questionnaire on their experience about the round they just played. This ordering was intentional as their subjective workload may change after performing the questionnaire. Once completed, facilitator F_A conducted a short interview on the game they just played and prompted the player on any extremities in their subjective ratings.

Player B (Naive Opponent) Procedure The procedure for the naive opponent (P_B) was straightforward, where the player was required to play three regular games against player P_A while being eye tracked, therefore acted as the control group. Once the players entered their respective rooms, the facilitator (F_B) calibrated the user to the eye tracker before the tutorial. At the end of each condition, we administered the NASA-TLX and a Games Experience Questionnaire (GEQ) [8]. The primary purpose of both questionnaires was for the player to fill up the time while player P_A went through a longer post-study questionnaire and the interview. Any gaps in time were filled up by facilitator F_B, who will engage in a conversation about the game they just played.

5.5   Results

The first part of this section presents the overall results from our various subjective and objective measures as outlined in Section 5.3. In the second, we present and discuss the experience of the players with and without the agent from insights provided by the post-study semi-structured interviews in relation to our various measures. Figure 4 summarises the median scores for the responses in our self-assessment questionnaire.

A Kruskal-Wallis test revealed no significant differences between the conditions for each of the measures. The figure shows that the conditions were found to be comparable except for the decreasing trend in distraction as we reduced the information. In addition to these measures, players in both AI conditions rated an agreeable median score for reliability (5.0) and when asked if they understood the AI predictions (6.0). The results suggest that although the communication was clear, the AI was unable to meet the expectation

of the player, such as by not providing correct predictions, predictions that the player already guessed or that the predictions were not timely enough for them to act on it.
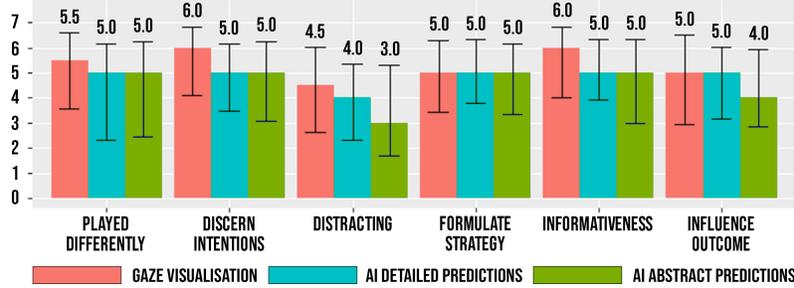

Fig. 4: Questionnaire Results.

Table 1 below shows the rating given for each condition in relation to *preference*, *demand* and *usefulness*. A Friedman test showed no significant differences between the conditions for all three ratings. These ratings, however, served as prompts for discussion during our interview as players reflected on their reasoning behind their given ratings.

Table 1: Post-Study Ratings for Each Condition

	Preferred			Demanding			Useful		
	Most	*Middle*	*Least*	*Most*	*Middle*	*Least*	*Most*	*Middle*	*Least*
GAZE VIZ	**15**	13	2	**15**	9	6	**13**	12	5
AI DETAILED	7	8	**15**	6	12	12	9	6	**15**
AI ABSTRACT	8	9	13	8	10	12	8	12	10

Table 2: Results of Cognitive Workload Measures.

	IPA (Hz)	Average Long Fixations (>500ms)	Average Saccade Velocity (m/sec)	NASA-TLX
GAZE VIZ	0.0127	0.02 (0.02)	47.80 (27.9)	49.1
AI DETAILED	0.0138	0.22 (0.10)	54.71 (29.2)	45.4
AI ABSTRACT	0.0148	0.02 (0.02)	49.49 (31.3)	45.9

Table 2 summarises the results of our cognitive workload measures. We ran a Mann-Whitney U test for all the objective measures and only found significant differences for the average long fixations measure. A post-hoc analysis showed differences between the GAZE VIZ and DETAILED AI PREDS conditions ($W=0$, $Z=4.78$, $p<0.05$, $r=0.87$) and between both AI conditions ($W=0$, $Z=4.78$, $p<0.05$, $r=0.87$). The results show that players on average had longer fixations in the DETAILED AI PREDS condition, which could simply be because participants needed time to process the predictions. We found no significant differences between GAZE VIZ and ABSTRACT AI PREDS conditions, which suggests that players did not require a longer time to parse the predictions in ABSTRACT AI PREDS condition, indicating that the ABSTRACT AI PREDS case did not introduce any significant burden on the users time but achieved as effective awareness as the GAZE VIZ condition. Though the AI did not decrease the cognitive workload as compared with the current approach of gaze visualisation, the NASA-TLX scores indicate that players perceived that GAZE VIZ condition to be more demanding overall than when being assisted by the agent.

However, although the overall mean score for the measure suggests the perceived workload for the gaze visualisation condition was higher when compared to the AI conditions, a Kruskal-Wallis test showed no significant differences among the three conditions.

AI Predictions Players who spoke positively about the predictions often referred to the specific properties in the predictions, including temporal and spatial properties as found in prior work (e.g. *"I like the temporal information ('since the beginning of the game...'), and precise information about where the opponent was looking."* – [P17$_A$]). The uncertainty provided in the explanations was also well received by players, noting that they only needed to know the areas than the specific cities (e.g. P8$_A$), or that the agent communicated alternate paths the opponent may take (e.g. P30$_A$). Player P12$_A$ explicitly noted that the predictions were useful when the agent predicted longer (distal) routes instead of shorter (proximal) routes, especially for strategy formulation.

There is some evidence to suggest that the AI predictions drew their attention to areas of the board they overlooked. For example, player P5$_A$ mentioned *"It made me take notice of what my opponent was doing."*. A third of players (10/30) noted that they had to invest time in deciphering the AI predictions, mostly attributed to their unfamiliarity with the map, despite each prediction having an overall indication of the area in the predictions where applicable (e.g. South East). This finding also brings forward an issue with the textual representation of intentions (*"I like the predictions that were short; I did not like the visuals. It was easier to take the AI info but not as pop-up prompts."* – [P29$_A$]; *"It took me out of the game a little to have the prompt pop up and then look at the map to interpret."* – [P10$_A$]). Player P1$_A$ mentions that *"...it would be better if the route was highlighted."* as a suggestion to complement the predictions with a concise visual component.

Players who least preferred the AI conditions found the prediction prompts distracting because it interrupted their thought process. As they were required to reflect on the prediction each time, it took them further away from their current task. Between the AI conditions, players preferred the abstract predictions over the detailed predictions in general as the information was more concise and therefore needed less time investment in understanding and utilising the information:

- *"I liked the short form prompts, they were actually quicker to read and I was still able to formulate a plan around my interpretation of the prompt."* – [P10$_A$]
- *"Shorter and brief hints were easy to understand and helpful."* – [P23$_A$]
- *"I liked the simplicity of the information it gave me, it was very easy to filter."* – [P8$_A$]

There were overarching reports that predictions became less useful as the game progressed as expected, especially towards the end of the game, as P12$_A$ mentions *"I liked the initial predictions but it was less helpful towards the end of the game"*. A possible explanation is because there was enough evidence in the form of routes claimed and players could make their own inferences through the observable opponent actions. As the agent lacked awareness of the context, players also noted several limitations in the AI conditions, such as not being able to predict whether the opponent was going to block them. (e.g. [P17$_A$]). Further, the agent was expected to communicate when prior predictions are no longer relevant (e.g. *"I'm not sure how helpful the AI was. It could be that the opponent did not have enough cards to carry out his original plan, or I blocked him successful at the beginning."* – [P15$_A$]). Furthermore, some players mentioned that they did not pay attention to their opponent's plans throughout the game, as their plans were not affected by their own. This finding suggests that the AI made them aware of their opponent's plans, but in some ways annoyed them as the AI kept informing them about the opponent's plans when it did not affect their plans throughout the game.

Gaze Visualisation A third of players (10/30) explicitly mentioned that the gaze visualisation was 'distracting', mentioning it *"moved too much"* [P1$_A$], occupied their time and attention [P15$_A$], which then caused them to play longer turns [P10$_A$]. When prompted further, three players (P12$_A$, P25$_A$, P29$_A$) mentioned it was mentally demanding to focus on their own and their opponent's strategies (or plans), causing a distraction.

Half the players (16/30) found the gaze visualisation to be informative and therefore useful, with a general consensus that it was good to know the general areas the opponent was looking at. Player P17$_A$ enjoyed the challenging aspect of inferring the opponent's intentions on his own, while P12$_A$ found it interesting to reaffirm his assumptions. Though these players found the visualisation informative, players also were not able to utilise the information that was available to them, especially if they were not experienced in the game (e.g. *"It was good to know the general areas the opponent was going for, but don't think I'm experienced enough to act well on the information."* – [P23$_A$]). These findings are also reflected in our questionnaire results as shown in Figure 4.

Table 1 shows that although gaze was found to be most demanding, it was rated most preferred and useful. There are two possible explanations for this. First, experienced players were able to utilise the additional information better through gaze. Second, players noted that the fact that the gaze was overlaid over the game made it easy to determine the areas of interest spatially, which was sufficient to gauge their opponents' intentions. A few players drew comparisons with the AI predictions, for example, player P23$_A$ mentions *"I prefer it [gaze visualisation] to the AI because I didn't have to bother with reading the pop-ups."*. Players can also mention that it was possible to ignore gaze when they want to, attributing it to visual background noise on the interface. However, players did note the ability to access the information at all times with gaze visualisation. In comparison with the AI conditions, the information was only available when the predictions appeared, leaving the players on occasion with no information when the opponent took a long turn.

6   Discussion

In this paper, we evaluated the prospects of an ideal intention-aware artificial agent, which we designed using goals derived from the literature. We present the first step towards artificial agents that can interpret and communicate intentions afforded by gaze input to assist a user by improving the user's situation awareness. Further, we evaluated whether the agent can alleviate the distracting nature of live gaze visualisation used to recognise intentions in prior work. To that end, we conducted two studies: first to derive a language model used by our agent to communicate the predicted intentions using natural language, and second to evaluate the effectiveness and experience of interacting with our agent.

The predictions and explanations provided by the agent early in the game allowed the participants to formulate better strategies, but overall the agent neither impacted the players' performance nor decreased the cognitive workload as initially hypothesised. It was possible that the game itself introduced cognitive workload, which is difficult to isolate as players had different abilities and set of goals. However, the overall perceived cognitive workload was lower in the agent-assisted conditions, with reduced distraction as compared to the gaze visualisation approach. We further acknowledge that irrespective of the mode of communication, the processing of information generates cognitive workload. Our subjective assessments indicate that the agent was successful in deriving intentions from gaze and communicating them to the participants in a way that matched the informativeness of the gaze visualisation. These results suggest that there is vast potential in using artificial agents to take on such roles when provided with complementary inputs

such as gaze. We also note that an agent-assisted approach can potentially scale well for multiple users, where the agent can determine what is the most relevant information to communicate, compared to visualising multiple user's gaze on the interface which would clutter the interface and potentially cause confusion. Due to the limitations of our approach concerning representation and context, we have only partially achieved our design goals for a collaborative intention-aware artificial agent. In the following subsections, we discuss the considerations when designing such agents from our findings.

Information Presentation A significant limitation of our approach is the full use of textual representations to convey human intentions. While this serves as a good starting point, it caused users in our study who were unfamiliar with the game to underutilise the predictions from the agent, as they needed to be spatially aware of the layout of the interface, i.e. the location of cities or map areas, to understand the predictions. Our findings suggest that an overlay of precise intentions over the interface (visual augmentation) by the agent coupled with natural language annotations can potentially be a more understandable way to communicate predictions and explanations.

Context-Awareness In our user study, we evaluated two sides of the interaction simultaneously. On one side, whether the agent can process and communicate intentions in real-time by observing a human player (the opponent), hence the *sender*. On other, the experience of the *receiver* of intentions, in our case, the agent-assisted player. Ideally, the agent should take into account what the agent-assisted player already knows by deriving their intentions as well, either implicitly or explicitly. With context-awareness, the agent would only communicate relevant predictions, such as predictions that directly affect the user, and which are therefore beneficial regardless of the mode of communication. The DETAILED AI PREDS condition was an extreme case where we gave the most complete explanation possible without considering what the participant already knew, leading to the communication of redundant information. The subjective assessment of this condition shows that it is indeed necessary to keep a model of what the player already knows, or what has already been communicated, to reduce distracting information and increase the effectiveness of each communication. Moreover, context-awareness would allow the agent to adjust the *level of detail* when communicating intentions. The combination of more concise information and more timely predictions would, we infer, improve the human's ability to respond to the agent. Furthermore, if the agent understood what all observable users intended to do, it would be possible for it to negotiate the goals of each of the users. Our work closely resembles *iTourist*, in which an agent could recognise gaze patterns of a 'tourist' and provide recommendations on transport or accommodation alternatives [52], but only for a single user at a time. We extend this work by demonstrating the ability of an automated system to understand long-term human intentions, and by providing insights on how these intentions can be communicated effectively, in a way that can be easily scaled to multiple users.

Nonverbal Communication This work provides an empirical assessment in a real-time setting of the intention prediction model that we developed in a previous paper [55], and shows that nonverbal inputs such as gaze can be used as a basis for natural language explanations. Further, this work demonstrates the usefulness of multimodal human inputs in the context of human-agent teaming. Our broader aim in this work is to provide a generalisable approach for designing such agents (we do not claim ecological validity for our study setting). Our work aims to improve on current approaches for human-awareness by not only detecting human presence or actions, but also predicting their

intentional actions. As an example, we use the work of Unhelkar et al. [57]'s human-robot collaborative assembly task. In their task, the work area was divided into cells, some shared by humans and robots, which were required to cease operating entirely whenever a human enters a shared cell. They developed and tested a model that incorporated predictions of human motion to improve the efficiency and safety of the assembly task. However, if the robot's motion planner could 'see' that as the human was moving towards the work area they consistently looked at a workbench in a cell that was different to their own, the robot could easily fuse the gaze and motion information to determine the cell that the human was going to and continue its work rather than stop. Such an approach we believe will further improve task efficiency and the interaction between the human and the robot. Hence, agents with the ability to process intentions can not only improve their interactions with their human counterparts but improve their proactiveness as well.

Explainable Agency Our first study formed the basis of a general model of intention communication, which can support the cognitive process of generating explanations involving observable actions and gaze behaviours. As explanations in an explainable agency [35, 40] involves both a cognitive process to derive an explanation and a social process of communicating the explanation to a human [42, 46], there is a clear scope of expanding our approach to generalise our findings to other settings, evaluate our existing approach [25], and to explore two-way communication between the human and the agent (e.g. dialogue). In essence, our agent posses the ability to keep a mental model of a users' short and long-term intentions, in which we can interrogate at any point in time. Lastly, our work focused on intention recognition aspect of explanation, which goes beyond question-answering, and differs from existing approaches where the presence of features is used to explain instead of the long-term observation of human behaviours.

7   Conclusions

In this paper, we have demonstrated a viable approach for designing the communication and interaction means for socially interactive agents, addressing prerequisites for effective human-agent collaboration [37, 38]. Our approach uses a proactive agent to assist a human player engaged in an online strategy game by improving the player's situation awareness through the communication of an opponent's intentions. We developed a language model based on human communication that allows our intention-aware agent to communicate inferred intentions through the observation of gaze behaviours and actions. In a user study, we evaluated the experience with and without the agent and found that players were receptive to the agent due to its ability to provide situation awareness of future intentions without the distractions of gaze visualisation. The agent's ability to digest gaze information into contextual and useful representations has broad implications for future systems. We provide several considerations on the design of such agents, in reference to the presentation of information, the need for context awareness, and opportunities in harnessing nonverbal communication. Overall, this paper highlights an opportunity for nonverbal behavioural inputs in Human-Agent Interaction, and provides an approach that can be used in scenarios where knowing the intentions of others is important (e.g. air traffic control, wargaming). In future work, we plan to extend the agent with the ability to consider additional input from the user, and generate alternative predictions about another user based on a 'what-if' query (such an action that the user is most likely to take). This will be particularly useful in collaborative scenarios, where an agent can assist, mediate or negotiate with knowledge of multiple users intentions.

Acknowledgements

References

1. Akkil, D., Dey, P., Salian, D., Rajput, N.: Gaze awareness in agent-based early-childhood learning application. In: Bernhaupt, R., Dalvi, G., Joshi, A., K. Balkrishan, D., O'Neill, J., Winckler, M. (eds.) Human-Computer Interaction - INTERACT 2017. pp. 447–466. Springer International Publishing, Cham (2017)
2. Akkil, D., Thankachan, B., Isokoski, P.: I see what you see: Gaze awareness in mobile video collaboration. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. pp. 32:1–32:9. ETRA '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3204493.3204542
3. Antifakos, S., Schwaninger, A., Schiele, B.: Evaluating the effects of displaying uncertainty in context-aware applications. In: Davies, N., Mynatt, E.D., Siio, I. (eds.) UbiComp 2004: Ubiquitous Computing. pp. 54–69. Springer Berlin Heidelberg, Berlin, Heidelberg (2004)
4. Argyle, M., Cook, M.: Gaze and mutual gaze. Cambridge U Press (1976)
5. 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)
6. Biran, O., Cotton, C.: Explanation and justification in machine learning: A survey. In: IJCAI-17 Workshop on Explainable AI (XAI). p. 8 (2017)
7. Brewer, J., D'Angelo, S., Gergle, D.: Iris: Gaze visualization design made easy. In: Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems. pp. D504:1–D504:4. CHI EA '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3170427.3186502
8. Brockmyer, J.H., Fox, C.M., Curtiss, K.A., McBroom, E., Burkhart, K.M., Pidruzny, J.N.: The development of the game engagement questionnaire: A measure of engagement in video game-playing. Journal of Experimental Social Psychology **45**(4), 624 – 634 (2009). https://doi.org/10.1016/j.jesp.2009.02.016
9. Buettner, R.: Cognitive workload of humans using artificial intelligence systems: Towards objective measurement applying eye-tracking technology. In: Timm, I.J., Thimm, M. (eds.) KI 2013: Advances in Artificial Intelligence. pp. 37–48. Springer Berlin Heidelberg, Berlin, Heidelberg (2013)
10. Chen, J., Barnes, M.: Human-agent teaming for multirobot control: A review of human factors issues. IEEE Transactions on Human-Machine Systems **44**(1), 13–29 (Feb 2014). https://doi.org/10.1109/THMS.2013.2293535
11. Chen, J.Y.C., Lakhmani, S.G., Stowers, K., Selkowitz, A.R., Wright, J.L., Barnes, M.: Situation awareness-based agent transparency and human-autonomy teaming effectiveness. Theoretical Issues in Ergonomics Science **19**(3), 259–282 (2018). https://doi.org/10.1080/1463922X.2017.1315750
12. D'Angelo, S., Begel, A.: Improving communication between pair programmers using shared gaze awareness. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. pp. 6245–6290. CHI '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/3025453.3025573

13. D'Angelo, S., Brewer, J., Gergle, D.: Iris: A tool for designing contextually relevant gaze visualizations. In: Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications. pp. 79:1–79:5. ETRA '19, ACM, New York, NY, USA (2019). https://doi.org/10.1145/3317958.3318228, http://doi.acm.org/10.1145/3317958.3318228

14. D'Angelo, S., Gergle, D.: Gazed and confused: Understanding and designing shared gaze for remote collaboration. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 2492–2496. CHI '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2858036.2858499

15. D'Angelo, S., Gergle, D.: An eye for design: Gaze visualizations for remote collaborative work. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 349:1–349:12. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3173923

16. 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). https://doi.org/10.1145/3173574.3174136

17. Duchowski, A.T.: Gaze-based interaction: A 30 year retrospective. Computers & Graphics **73**, 59 – 69 (2018). https://doi.org/10.1016/j.cag.2018.04.002

18. Duchowski, A.T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., Raubal, M., Giannopoulos, I.: The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 282:1–282:13. CHI '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3173856

19. 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, New York, NY, USA (2018). https://doi.org/10.1145/3172944.3172961

20. Endsley, M.R.: Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors: The Journal of the Human Factors and Ergonomics Society **37**(1), 32–64 (1995). https://doi.org/10.1518/001872095779049543

21. Harbers, M., van den Bosch, K., Meyer, J.J.C.: A study into preferred explanations of virtual agent behavior. In: Ruttkay, Z., Kipp, M., Nijholt, A., Vilhjalmsson, H.H. (eds.) Intelligent Virtual Agents. pp. 132–145. Springer Berlin Heidelberg, Berlin, Heidelberg (2009)

22. Hart, S.G., Staveland, L.E.: Development of nasa-tlx (task load index): Results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (eds.) Human Mental Workload, Advances in Psychology, vol. 52, pp. 139 – 183. North-Holland (1988). https://doi.org/https://doi.org/10.1016/S0166-4115(08)62386-9

23. Hayes, B., Shah, J.A.: Improving robot controller transparency through autonomous policy explanation. In: Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction. pp. 303–312. HRI '17, ACM, New York, NY, USA (2017). https://doi.org/10.1145/2909824.3020233, http://doi.acm.org/10.1145/2909824.3020233

24. Higuch, K., Yonetani, R., Sato, Y.: Can eye help you?: Effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. pp. 5180–5190. CHI '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2858036.2858438

25. Hoffman, R.R., Mueller, S.T., Klein, G., Litman, J.: Metrics for explainable ai: Challenges and prospects. arXiv preprint arXiv:1812.04608 (2018)

26. Huang, C.M., Andrist, S., Sauppé, A., Mutlu, B.: Using gaze patterns to predict task intent in collaboration. Frontiers in Psychology **6**, 1049 (2015). https://doi.org/10.3389/fpsyg.2015.01049

27. Huang, C.M., Mutlu, B.: Anticipatory robot control for efficient human-robot collaboration. In: The Eleventh ACM/IEEE International Conference on Human

Robot Interaction. pp. 83–90. HRI '16, IEEE Press, Piscataway, NJ, USA (2016), http://dl.acm.org/citation.cfm?id=2906831.2906846

28. 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). https://doi.org/10.1145/97243.97246

29. Kahneman, D., Egan, P.: Thinking, fast and slow, vol. 1. Farrar, Straus and Giroux New York (2011)

30. Kass, R., Finin, T., et al.: The need for user models in generating expert system explanations. International Journal of Expert Systems **1**(4) (1988)

31. Keil, F.C.: Explanation and understanding. Annual Review of Psychology **57**(1), 227–254 (2006). https://doi.org/10.1146/annurev.psych.57.102904.190100

32. Kelley, J.F.: An iterative design methodology for user-friendly natural language office information applications. ACM Trans. Inf. Syst. **2**(1), 26–41 (Jan 1984)

33. Klien, G., Woods, D.D., Bradshaw, J.M., Hoffman, R.R., Feltovich, P.J.: Ten challenges for making automation a "team player" in joint human-agent activity. IEEE Intelligent Systems **19**(6), 91–95 (Nov 2004). https://doi.org/10.1109/MIS.2004.74

34. Land, M.F.: Vision, eye movements, and natural behavior. Visual Neuroscience **26**(1), 51–62 (2009). https://doi.org/10.1017/S0952523808080899

35. Langley, P., Meadows, B., Sridharan, M., Choi, D.: Explainable agency for intelligent autonomous systems. In: Twenty-Ninth IAAI Conference (2017)

36. Lankes, M., Maurer, B., Stiglbauer, B.: An eye for an eye: Gaze input in competitive online games and its effects on social presence. In: Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology. pp. 17:1–17:9. ACE2016, ACM, New York, NY, USA (2016). https://doi.org/10.1145/3001773.3001774

37. Lesh, N., Marks, J., Rich, C., Sidner, C.L.: Man-computer symbiosis revisited: Achieving natural communication and collaboration with computers. IEICE Transactions **87**(6), 1290–1298 (2004)

38. Licklider, J.C.R.: Man-computer symbiosis. IRE Transactions on Human Factors in Electronics **HFE-1**(1), 4–11 (March 1960). https://doi.org/10.1109/THFE2.1960.4503259

39. Lim, B.Y., Dey, A.K.: Investigating intelligibility for uncertain context-aware applications. In: Proceedings of the 13th International Conference on Ubiquitous Computing. pp. 415–424. UbiComp '11, ACM, New York, NY, USA (2011). https://doi.org/10.1145/2030112.2030168

40. Madumal, P.: Explainable agency in intelligent agents: Doctoral consortium. In: Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems. pp. 2432–2434. AAMAS '19, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2019), http://dl.acm.org/citation.cfm?id=3306127.3332137

41. Madumal, P., Miller, T., Sonenberg, L., Vetere, F.: Explainable reinforcement learning through a causal lens. arXiv preprint arXiv:1905.10958 (2019)

42. 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

43. Majaranta, P., Bulling, A.: Eye Tracking and Eye-Based Human–Computer Interaction, pp. 39–65. Springer London, London (2014)

44. 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)

45. Maurer, B., Lankes, M., Stiglbauer, B., Tscheligi, M.: EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game, pp. 102–114. Springer International Publishing, Cham (2016)

46. Miller, T.: Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence **267**, 1 – 38 (2019). https://doi.org/https://doi.org/10.1016/j.artint.2018.07.007

47. Moore, J.D., Paris, C.L.: Requirements for an expert system explanation facility. Computational Intelligence **7**(4), 367–370 (1991)
48. Narayanan, V., Zhang, Y., Mendoza, N., Kambhampati, S.: Automated planning for peer-to-peer teaming and its evaluation in remote human-robot interaction. In: Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts. pp. 161–162. HRI'15 Extended Abstracts, ACM, New York, NY, USA (2015). https://doi.org/10.1145/2701973.2702042
49. 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
50. Newn, J., Velloso, E., Allison, F., Abdelrahman, Y., Vetere, F.: Evaluating real-time 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
51. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. User Modeling and User-Adapted Interaction **27**(3-5), 393–444 (Dec 2017). https://doi.org/10.1007/s11257-017-9195-0
52. Qvarfordt, P., Zhai, S.: Conversing with the user based on eye-gaze patterns. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 221–230. CHI '05, ACM, New York, NY, USA (2005). https://doi.org/10.1145/1054972.1055004
53. Riek, L.D.: Wizard of oz studies in hri: A systematic review and new reporting guidelines. J. Hum.-Robot Interact. **1**(1), 119–136 (Jul 2012). https://doi.org/10.5898/JHRI.1.1.Riek
54. Rieser, V.: Bootstrapping reinforcement learning-based dialogue strategies from wizard-of-oz data. DFKI (2008)
55. 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 System. AAMAS '18, International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC (2018), http://dl.acm.org/citation.cfm?id=3237383.3237457
56. Stein, R., Brennan, S.E.: Another person's eye gaze as a cue in solving programming problems. In: Proceedings of the 6th International Conference on Multimodal Interfaces. pp. 9–15. ICMI '04, ACM, New York, NY, USA (2004). https://doi.org/10.1145/1027933.1027936
57. Unhelkar, V.V., Lasota, P.A., Tyroller, Q., Buhai, R., Marceau, L., Deml, B., Shah, J.A.: Human-aware robotic assistant for collaborative assembly: Integrating human motion prediction with planning in time. IEEE Robotics and Automation Letters **3**(3), 2394–2401 (July 2018). https://doi.org/10.1109/LRA.2018.2812906
58. Velloso, E., Carter, M.: The emergence of eyeplay: A survey of eye interaction in games. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. pp. 171–185. CHI PLAY '16, ACM, New York, NY, USA (2016). https://doi.org/10.1145/2967934.2968084, http://doi.acm.org/10.1145/2967934.2968084
59. Vertegaal, R.: The gaze groupware system: Mediating joint attention in multiparty communication and collaboration. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 294–301. CHI '99, ACM, New York, NY, USA (1999). https://doi.org/10.1145/302979.303065, http://doi.acm.org/10.1145/302979.303065
60. Wang, D., Yang, Q., Abdul, A., Lim, B.Y.: Designing theory-driven user-centric explainable ai. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 601:1–601:15. CHI '19, ACM, New York, NY, USA (2019)
61. Wang, N., Pynadath, D.V., Hill, S.G.: Trust calibration within a human-robot team: Comparing automatically generated explanations. In: The Eleventh ACM/IEEE International Conference on Human Robot Interaction. pp. 109–116. IEEE Press (2016)
62. Zhang, Y., Pfeuffer, K., Chong, M.K., Alexander, J., Bulling, A., Gellersen, H.: Look together: using gaze for assisting co-located collaborative search. Personal and Ubiquitous Computing **21**(1), 173–186 (Feb 2017). https://doi.org/10.1007/s00779-016-0969-x