Gaze-Based Intention Recognition for Human-Agent Collaboration

Towards Nonverbal Communication in Human-AI Interaction

Joshua Newn
(https://orcid.org/0000-0001-5769-6297)

Submitted in total fulfilment of the requirements of the degree of

Doctor of Philosophy

School of Computing and Information Systems
The University of Melbourne

March 2020

Copyright © 2020 Joshua Newn

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm or any other means without written permission from the author.

Abstract

Human-agent collaboration has repeatedly been proposed over the decades as a way forward to leverage the strengths of artificial intelligence. As it has become common for humans to work and play alongside intelligent agents, it is increasingly imperative to improve the capacity of agents to interact with their human counterparts socially, naturally and effectively. However, current agents are still limited in their capacity to recognise nonverbal signals and cues, which in turn, limits their capabilities for natural interaction. This thesis addresses this limitation by investigating how artificial agents might support humans in real-time collaboration, given the increased capacity to recognise human intentions afforded by processing their gaze data in real-time. We hypothesise that a socially interactive agent with an increased capacity to recognise intentions can drastically improve its interactive capability, such as by adapting its recommendations to their anticipated intentions as well as to the intentions of others.

Using a scenario-based based design approach, we designed five studies to inform and evaluate the different capabilities of a collaborative gaze-enabled intention-aware artificial agent.

In Studies 1 and 2, we first evaluate the capacity of human subjects to perform intention recognition using gaze visualisation and its corresponding effects in a competitive gameplay setting. The findings showed that humans players could improve their capacity to infer their intentions of their opponent when shown a live visualisation of their opponent's gaze throughout the game. However, this capacity can be hampered when the opponent was aware that their gaze was being watched. The findings further indicate that humans have a limited capacity in performing gaze-based intention recognition, suggesting that the task may be more suitable for an artificial agent that is trained to process the rich multimodal information available in our setting.

In Study 3, we present the implementation details and evaluation of a gaze-enabled intention-aware artificial agent, developed as part of this thesis, that incorporates gaze into its intention recognition process. The evaluation, which uses the data from the previous two studies, demonstrates that incorporating gaze into the agent's planning process not only increases the agent's capacity to recognise intentions but also that it performs better overall than human subjects. In Studies 4 and 5, we operationalise the artificial agent by first giving the agent both the ability to communicate intentions of their opponent to its human collaborator and to explain its reasoning process if required. Subsequently, we evaluated the experience of the players playing the game with and without the assistance of the agent, which ultimately provided insights into how we can further improve the interaction between the human and an intention-aware artificial agent.

The findings in this thesis resulted in three contributions towards the understanding of how artificial agents can support human-agent collaboration, given the ability increased capacity to recognise intentions with eye-tracking. The findings from Studies 1, 2 and 3 extend the relationship between gaze awareness and intention, by demonstrating that gaze when tracked over time, can lead to the detection of distal intentions (i.e. long-term intentions that often require several steps to be fulfilled). Following, Studies 3, 4 and 5 contribute to the design of a collaborative gaze-enabled intention-aware artificial agent, and the demonstration of increased situation awareness through gaze awareness for human-agent collaboration. Overall, the thesis highlights the importance of incorporating nonverbal communication in human-AI interaction.

Declaration

I, Joshua Newn, declare that this thesis titled, "Gaze-Based Intention Recognition for Human-Agent Collaboration: Towards Nonverbal Communication in Human-AI Interaction" and the work presented in it are my own. I confirm that

1. the thesis comprises only the candidate's original work towards the degree of Doctor of Philosophy, except where indicated in the preface;
2. due acknowledgement has been made in the text to all other material used; and
3. the thesis is fewer than the 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Signed: *Joshua Newn*

Date: *20/03/2020*

Preface

This thesis includes three published papers, in accordance with The University of Melbourne's guidelines for a *Thesis with Publication*[1]. For all three papers, my thesis supervisors, Prof Frank Vetere and Dr Eduardo Velloso contributed feedback on the study design and written drafts.

Chapter 4 presents the following peer-reviewed conference paper:

Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. 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* (CHI PLAY '17). Association for Computing Machinery, New York, NY, USA, 541-552. DOI: https://doi.org/10.1145/3116595.3116624

As the first author, I was responsible for the design of the study, the literature review, data collection and analysis, and preparation for the work for publication under the supervision of my supervisors. I drafted the work and performed subsequent editing in response to reviews from co-authors and reviewers. Co-authors Fraser Allison (fellow PhD Candidate) and Yomna Abdelrahman (visiting PhD Candidate from the University of Stuttgart) assisted with the data analysis and write up of the findings. The publication is not subject to any obligation or contractual agreements with a third party that would constrain its inclusion in this thesis.

[1] https://gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

Chapter 5 presents a published peer-reviewed conference paper:

Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. 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 (CHI '18). Association for Computing Machinery, New York, NY, USA, Paper 261, 1–12. DOI: https://doi.org/10.1145/3173574.3173835

As the first author, I was responsible for the design of the study, the literature review, data collection and analysis, and preparation for the work for publication under the supervision of my supervisors. I drafted the work and performed subsequent editing in response to reviews from co-authors and reviewers. Co-author Fraser Allison (fellow PhD Candidate) assisted with the study as the co-experimenter, qualitative data analysis and write up of the findings. The publication is not subject to any obligation or contractual agreements with a third party that would constrain its inclusion in this thesis.

Chapter 7 and 8 presents a published peer-reviewed conference paper:

Joshua Newn, Ronal Singh, Fraser Allison, Prashan Madumal, Eduardo Velloso, and Frank Vetere. 2019. Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents. In *17th IFIP International Conference on Human-Computer Interaction — INTERACT 2019* (INTERACT '19). Lecture Notes in Computer Science, 11747. Springer, Cham. DOI: https://doi.org/10.1007/978-3-030-29384-0_17

As the first author, I was responsible for the design of the study, the literature review, data collection and analysis, and preparation for the work for publication under the supervision of my supervisors. I drafted the work and performed subsequent editing in response to reviews from co-authors and reviewers. Co-authors Ronal Singh, Fraser Allison and Prashan Madumal were co-experiments in the study, and further assisted with the data analysis and write up of the findings. Ronal Singh assisted with the technical implementation and testing of the apparatus employed in the evaluation study. The publication is not subject to any obligation or contractual agreements with a third party that would constrain its inclusion in this thesis.

The following peer-reviewed papers were produced during my PhD candidature that drew upon the research conducted for this thesis:

Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. 2016. Exploring the Effects of Gaze Awareness on Multiplayer Gameplay. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (CHI PLAY Companion '16). Association for Computing Machinery, New York, NY, USA, 239-244. DOI: https://doi.org/10.1145/2968120.2987740

Joshua Newn. 2018. Enabling Intent Recognition Through Gaze Awareness in User Interfaces. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems* (CHI EA '18). Association for Computing Machinery, New York, NY, USA, Paper DC15, 4 pages. DOI: https://doi.org/10.1145/3170427.3173028

Ronal Singh, Tim Miller, **Joshua Newn**, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. 2018. Combining Planning with Gaze for Online Human Intention Recognition. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 488-496. URL: http://dl.acm.org/citation.cfm?id=3237383.3237457

Michael Lankes, **Joshua Newn**, Bernhard Maurer, Eduardo Velloso, Martin Dechant, and Hans Gellersen. 2018. EyePlay Revisited: Past, Present and Future Challenges for Eye-Based Interaction in Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (CHI PLAY '18 Extended Abstracts). Association for Computing Machinery, New York, NY, USA, 689-693. DOI: https://doi.org/10.1145/3270316.3271549

Prashan Madumal, Ronal Singh, **Joshua Newn**, and Frank Vetere. 2018. Interaction Design for Explainable AI: Workshop Proposal. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction* (OzCHI '18). Association for Computing Machinery, New York, NY, USA, 607-608. DOI: https://doi.org/10.1145/3292147.3293450

Joshua Newn, Ronal Singh, Eduardo Velloso and Frank Vetere. 2019. Combining Implicit Gaze and AI for Real-Time Intention Projection. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 324-327. DOI: https://doi.org/10.1145/3341162.3343786

Joshua Newn, Benjamin Tag, Ronal Singh, Eduardo Velloso, and Frank Vetere. 2019. AI-Mediated Gaze-Based Intention Recognition for Smart Eyewear: Opportunities & Challenges. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 637-642. DOI: https://doi.org/10.1145/3341162.3348387

Ronal Singh, Tim Miller, **Joshua Newn**, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. 2020. Combining Gaze and AI Planning for Online Human Intention Recognition. In *Artificial Intelligence*, 248, 103275:1–103275:26. DOI: https://doi.org/10.1145/3170427.3173028

The following peer-reviewed papers were produced during my PhD candidature, but did not draw upon or contribute directly to the research conducted for this thesis:

Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. 2016. Multimodal Segmentation on a Large Interactive Tabletop: Extending Interaction on Horizontal Surfaces with Gaze. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces* (ISS '16). Association for Computing Machinery, New York, NY, USA, 251–260. DOI: https://doi.org/10.1145/2992154.2992179

Eduardo Velloso, Marcus Carter, **Joshua Newn**, Augusto Esteves, Christopher Clarke, and Hans Gellersen. 2017. Motion Correlation: Selecting Objects by Matching Their Movement. In *ACM Transactions on Computer-Human Interaction* (TOCHI). 24, 3, Article 22 (July 2017), 35 pages. DOI: https://doi.org/10.1145/3064937

Oludamilare Matthews, Zhanna Sarsenbayeva, Weiwei Jiang, **Joshua Newn**, Eduardo Velloso, Sarah Clinch, and Jorge Goncalves. 2018. Inferring the Mood of a Community From Their Walking Speed: A Preliminary Study. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (UbiComp '18). Association for Computing Machinery, New York, NY, USA, 1144–1149. DOI: https://doi.org/10.1145/3267305.3274759

Namrata Srivastava, **Joshua Newn**, and Eduardo Velloso. 2018. Combining Low and Mid-Level Gaze Features for Desktop Activity Recognition. In *Proceedings of ACM Interactive, Mobile, Wearable and Ubiquitous Technologies* (IMWUT). 2, 4, Article 189 (December 2018), 27 pages. DOI: https://doi.org/10.1145/3287067

Fraser Allison, **Joshua Newn**, Wally Smith, Marcus Carter, and Martin Gibbs. 2019. Frame Analysis of Voice Interaction Gameplay. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). Association for Computing Machinery, New York, NY, USA, Paper 393, 1–14. DOI: https://doi.org/10.1145/3290605.3300623

Qiushi Zhou, **Joshua Newn**, Namrata Srivastava, Tilman Dingler, Jorge Goncalves, and Eduardo Velloso. 2019. Cognitive Aid: Task Assistance Based On Mental Workload Estimation. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19). Association for Computing Machinery, New York, NY, USA, Paper LBW2315, 1–6. DOI: https://doi.org/10.1145/3290607.3313010

Niels Wouters, Ryan Kelly, Eduardo Velloso, Katrin Wolf, Hasan Shahid Ferdous, **Joshua Newn**, Zaher Joukhadar, and Frank Vetere. 2019. Biometric Mirror: Exploring Ethical Opinions towards Facial Analysis and Automated Decision-Making. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (DIS '19). Association for Computing Machinery, New York, NY, USA, 447–461. DOI: https://doi.org/10.1145/3322276.3322304

Yomna Abdelrahman, Anam Ahmad Khan, **Joshua Newn**, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. 2019. Classifying Attention Types with Thermal Imaging and Eye Tracking. In *Proceedings of ACM Interactive, Mobile, Wearable and Ubiquitous Technologies* (IMWUT). 3, 3, Article 69 (September 2019), 27 pages. DOI: https://doi.org/10.1145/3351227

Qiushi Zhou, **Joshua Newn**, Benjamin Tag, Hao-Ping Lee, Chaofan Wang, and Eduardo Velloso. 2019. Ubiquitous smart eyewear interactions using implicit sensing and unobtrusive information output. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers* (UbiComp/ISWC '19 Adjunct). Association for Computing Machinery, New York, NY, USA, 661–666. DOI: https://doi.org/10.1145/3341162.3348392

Ebrahim Babaei, Namrata Srivastava, **Joshua Newn**, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. 2020. Faces of Focus: A Study on the Facial Cues of Attentional States. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20). Association for Computing Machinery, New York, NY, USA, 1–13. DOI: https://doi.org/10.1145/3313831.3376566

Ethics approval to conduct the studies comprising this thesis was provided by The University of Melbourne's Human Ethics Advisory Group, Engineering (Ethics ID: 1647972).

Acknowledgements

First and foremost, I would like to thank my supervisors, Prof Frank Vetere and Dr Eduardo Velloso, for once again guiding me through the treacherous waters of graduate research. I could not have picked a better pair of supervisors, and words cannot express the amount of faith, patience, encouragement and support that they have given me over the years. Thank you for believing in me and for embarking with me on this unforgettable once in a lifetime journey.

Frank, thank you for always being supportive in my career development as a researcher and for always looking out for opportunities that might suit me. I was a young, naive and inexperienced student when we met that one fine day in 2014 — looking back at that moment, who knew that I would be completing a *Doctor of Philosophy* under your supervision years later.

Eduardo, thank you for your continued mentorship and for taking the time to show me the ropes to succeed in academia. I always enjoyed our long chats, whether its over a coffee or a beer. Till today, it is great knowing that you still have my back and that you are always looking out for me. Most of all, it is great knowing I made a friend that I can to turn to in times of need.

Thank you to all my collaborators over the years, it was a pleasure working with all of you. To comrades Fraser Allison, Ronal Singh and Yomna Abdelrahman in particular, this thesis would not be what it is without your contributions. Fraser, thank you for being on-board across all three human-centred studies. I could not have picked a better co-experimenter and co-author. Ronal, thank you for the endless hours you have put into our collaborations. I enjoyed our long conversations about work and life. Yomna, thank you for being there to motivate me over the years. I learnt a great deal from our friendship, and I especially enjoy our train rides. Also thank you for hosting me not once but twice in Germany, and I will not forget your generosity.

I acknowledge Marcus Carter, who was on my original supervision team and also mentor in my early days as a young researcher. I learnt a lot from you from my first research project, which led to my very first publication. Thank you for your continued encouragement and support.

A special thank you to many others for making my PhD candidature a memorable experience. To Tim Miller and Melissa Rogerson, for keeping me afloat towards the end of my PhD and for many months after. It was the lifeline I needed. To my wonderful neighbours, Deepti Aggarwal and Rohit Khot, you have both been such wonderful neighbours, and it was indeed a delight sharing a wall with two fellow HCI researchers. To Zaher Joukhadar and Nicole Barbee, for your help over the years. To Niels van Berkel and Namrata Srivastava, for your everlasting support and encouragement, especially when times got tough for me. To Vassilis Kostakos, Jorge Goncalves and Tilman Dingler for involving me in various social and research activities. To all my peers, especially those who were part of Social NUI and the Interaction Design Lab (Sarah Webber, Romina Carrasco, Travis Cox, Ryan Kelly, Yousef Kowsar, Kagonya Awori, Unni Krishnan, Fernando Estrada, Kayla Heffernan, Jane Mavoa, Martin Reinoso, Prashan Madumal, Hasan Ferdous, and many more), for making the group feel like a big family. To my wonderful 'pod of six' (and across) in Room 5.10 (Danula Hettiachchi, Zhanna Sarsenbayeva, Chaofan Wang, Gabriele Marini, Kangning Yang, Qiushi Zhou and Brandon Syiem) for all the food, random conversations, trips and fun times. To all those who have worked with me in various capacities (too many to name), thank you for your trust as it gave me the confidence to perform at my very best. Lastly, to all participants involved, as this thesis would not have been possible without you.

I further acknowledge the Australian Commonwealth Government, the School of Computing and Information Systems, the Microsoft Research Centre for Social NUI, the Melbourne School of Engineering and Google for supporting the research projects conducted in the thesis.

A big thank you to my family, who has been proud of my achievements no matter how small and to have raised me well. To my mum and dad especially, who worked hard so that I can have the best things in life. I have always done what I wanted, and I am grateful that my decisions have always been supported by the both of you, no matter how delirious they may seem.

Finally, thank you to Jennifer Mulyawan from the bottom of my heart. I am extremely lucky to have you by my side and to have received your unconditional love and support. You helped me get up when I was at my lowest, and I would not have completed this thesis without you.

I dedicate this thesis to the memory of two individuals that departed during my PhD candidature.

To my grandfather, Augustine Bartholomew Newn Shak Hee (1927-2018), who provided me with the comforts and foundations in life to pursue knowledge. Kung, I hope that you are proud that the first Dr. Newn is from your lineage.

To Huyen Thi Thanh Nguyen (1992-2019), who inspired and supported me to take on this journey; who taught me to live life to the fullest with no regrets; who made me believe that anything is possible. May her gentle soul rest in peace.

Contents

List of Figures **xviii**

List of Tables **xx**

1 Introduction **1**

1.1 Background 5
1.2 Research Opportunity 7
1.3 Aim and Scope 10
1.4 Thesis Contributions 10
1.5 Thesis Overview 12

2 Literature Review **19**

2.1 Introduction 19
2.2 Human-Agent Collaboration 20
2.2.1 Overview 20
2.2.2 Utilisation of Nonverbal Communication Modalities 23
2.2.3 Matching the Language Between the Counterparts 24
2.2.4 Ability to Represent the Mental States of the User 25
2.2.5 Complementary and Effective Division of Labour 26
2.3 The Role of Gaze Awareness in Human Collaboration 27
2.4 The Role of Gaze Awareness in Human-Computer Interaction 29
2.4.1 Gaze Visualisation in Computer-Mediated Human-Human Interaction . 30
2.4.2 Gaze-Assisted Human-Computer Interaction 33
2.4.3 Gaze in Human-Robot Interaction 34
2.5 Gaze-Based Intention Recognition 35
2.6 Situation Awareness in Human-Agent Teaming 37

2.7 Research Questions . . . 38

3 Research Design 41

3.1 Introduction . . . 41
3.2 Research Approach: Scenario-Based Design . . . 42
3.3 Ticket to Ride . . . 47
3.3.1 Gameplay and Rules . . . 48
3.3.2 Concept . . . 50
3.3.3 Feasibility Study . . . 51
3.4 Structure of Studies . . . 54

4 Study 1: Real-Time Gaze Visualisations for Human-Human Intention Recognition 73

4.1 Introduction . . . 74
4.2 Pilot Study . . . 77
4.3 Published Paper . . . 84

5 Study 2: Effect of Gaze Awareness on Human-Human Intention Recognition 110

5.1 Introduction . . . 111
5.2 Published Paper . . . 115
5.3 Addendum: Further Analysis and Discussion on Deception . . . 142
5.4 Chapter Summary . . . 145

6 Study 3: Intention-Aware Gaze-Enabled Artificial Agent Evaluation 146

6.1 Introduction . . . 147
6.2 Related Work and Approach . . . 150
6.3 Model . . . 152
6.3.1 Problem Formulation . . . 152
6.3.2 Intention Recognition . . . 154
6.4 Model Evaluation . . . 159
6.4.1 Data Set . . . 159
6.4.2 Gaze Features and Planning . . . 160
6.4.3 Evaluation Setup and Performance Measures . . . 161

6.4.4 Technical Implementation . . . 163
6.5 Results . . . 164
6.5.1 Performance Comparison Between *Model Only* with *Gaze+Model* . . . 164
6.5.2 Performance Comparison Between *Human Players* and *Artificial Agent* . 168
6.6 Chapter Summary and Conclusion . . . 169

7 Study 4: Assistive Intention-Aware Artificial Agent Language Identification 170
7.1 Introduction . . . 171
7.2 Related Work and Approach . . . 174
7.3 Study . . . 176
7.3.1 Participants . . . 176
7.3.2 Experimental Setup and Procedure . . . 176
7.4 Study Findings . . . 179
7.4.1 Prediction Format . . . 179
7.4.2 Explanation Content . . . 180
7.4.3 Reasoning and Communication Strategies . . . 181
7.5 Chapter Discussion and Conclusion . . . 182

8 Study 5: Assistive Intention-Aware Artificial Agent Real-Time Interaction Evaluation 183
8.1 Introduction . . . 184
8.2 Study . . . 187
8.2.1 Intention-Aware Gaze-Enabled Artificial Agent . . . 187
8.2.2 Experimental Setup . . . 188
8.2.3 Study Conditions . . . 189
8.2.4 Measures and Analysis . . . 190
8.2.5 Participants . . . 191
8.2.6 Procedures . . . 191
8.3 Results . . . 193
8.3.1 AI Predictions . . . 195
8.3.2 Gaze Visualisation . . . 197
8.4 Discussion . . . 198

8.5 Conclusion . . . 201

9 General Discussion **203**

9.1 Introduction . . . 203

9.2 Research Questions Revisited . . . 208

9.3 Key Contributions and Implications . . . 211

9.3.1 Extending the Relationship Between Gaze Awareness and Intention . . . 211

9.3.2 Design of a Collaborative Intention-Aware Artificial Agent . . . 213

9.3.3 Demonstration of Gaze Awareness for Human-Agent Collaboration . . . 214

9.3.4 Implications for Gaze in Human-Agent Collaboration . . . 215

9.3.5 Implications for Gaze in Human-AI Interaction . . . 216

9.4 Limitations and Future Work . . . 218

9.4.1 Artificial Agent Intention Recognition Performance . . . 218

9.4.2 Artificial Agent Communication Aspects . . . 219

9.4.3 Artificial Agent Collaboration Capacity . . . 223

9.5 Concluding Remarks . . . 225

Appendix A Published Papers

A.1 Paper: *Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games* . . .

A.2 Paper: *Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay* . . .

A.3 Paper: *Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents*

Appendix B Study Materials

B.1 Study 2 Study Materials: Questionnaires . . .

B.2 Study 5 Study Materials: Questionnaires . . .

List of Figures

1.1 Thesis Research Structure. 13

3.1 Envisioned Human-Agent Collaboration Scenario – *Collaboration in a Competitive Setting.* 44

3.2 Ticket to Ride Board Game (and its areas of interest and components). 48

3.3 Gaze-Based Intention Recognition Concept (in relation to *Ticket to Ride*). 51

3.4 Feasibility Study Experimental Setup. 52

3.5 Example of early planning behaviour shown using gaze visualisation. 53

3.6 Example of how players track the gaze of one another during the game. 54

3.7 Breakdown of overarching envisioned scenario into *interaction-level scenarios.* . . 55

3.8 Overall Structure of Thesis Study Chapters. 58

3.9 Relationship between human-centred prerequisites, studies and outcomes. . . . 70

4.2 Screenshot of the digital desktop version *Ticket to Ride* board game. 77

4.3 Gaze visualisation example (using a red dot for visualisation). 78

4.4 Pilot study conditions. The circles represent a gaze visualisation overlay on the screen. 79

4.5 Gaze Visualisation System Architecture. 80

4.6 An opponent's gaze represented as a real-time heatmap visualisation over the *Ticket to Ride* game, showing interest in the east coast of the North America. . . 85

4.7 Real-Time Gaze Representations Evaluated. 89

4.8 Study 1 Study Setting 98

4.9 Study 1 Analysis Methods 100

4.10 Preference Ratings 104

5.2 Study 2 Conditions 122

5.3 Ticket to Ride Gaze Visualisation Prediction Example 130

5.4 Deception Analysis. Comparison of fixation duration in seconds and fixations counts on ground truth for the natural gaze (CU-GU) condition and the deceptive gaze condition (CA-GU). . . . 142
5.5 Comparison of gaze features between the natural and deceptive conditions. . . . 144

6.3 Visualisation of Artificial Agent Implementation. . . . 163
6.4 Agent Performance Example . . . 166
6.5 Visualisation of artificial agent performing early prediction using with gaze. . . . 167

7.2 Relationship between the *human-centred prerequisites*, studies and outcomes. . . . 173
7.3 Study 4 'Wizard of Oz' Experimental Setup. . . . 177
7.4 Study 4 Participant View. . . . 177

8.1 The figure illustrates how the outcomes of the previous studies converge in Study 5, where we operationalise the artificial agent in our chosen human-agent collaboration scenario. . . . 184
8.2 The outcomes of each study in relation to the *human-centred prerequisites*. . . . 186
8.3 Study 5 Experimental Setup. . . . 189
8.4 Study 5 Study Conditions . . . 189
8.5 Repeated-Measures Questionnaire Results . . . 193

9.1 Overview of thesis studies and outcomes (and how they related to each other). . . 205
9.2 Relationship between human-centred prerequisites for man-computer symbiosis and the studies conducted in this thesis. . . . 206
9.3 Example of visual explanation over the Ticket to Ride game. . . . 220

List of Tables

1.1 Summary of prerequisites for the realisation of Licklider's *Man-Computer Symbiosis*. 5

4.1 Real-time gaze representations in relation to conspicuousness and visual information. . . . 90

4.2 AOI Estimation Results. Precision (Recall). . . . 101

4.3 Route Prediction Results. Precision (Recall). . . . 102

4.4 Ratings of Informativeness Results 1 Not Informative at all - 7 Extremely Informative). Score (Standard Deviation) . . . 103

4.5 Predictive Ability Results. Score (Standard Deviation) . . . 103

5.1 Prediction Performance Results . . . 128

6.1 Proximal intention prediction for the two approaches. . . . 165

6.2 Performance comparison between human players and artificial agent for the natural and deceptive conditions. . . . 168

7.1 Key Prediction Formats . . . 180

8.1 Post-Study Ratings for Each Condition . . . 194

8.2 Results of Cognitive Workload Measures. . . . 194

The hope is that, in not too many years, human brains and computing machines will be coupled together very tightly, and that the resulting partnership will think as no human brain has ever thought and process data in a way not approached by the information-handling machines we know today.

– J.C.R. Licklider, *Man-Computer Symbiosis (1960)*

1 Introduction

This thesis explores the importance of eye gaze in the context of human-agent collaboration, focusing on its ability to provide implicit cues to improve the capacity of socially interactive artificial agents to recognise the intentions of their human counterparts. As it is now commonplace for humans to work and play alongside intelligent artificial agents, it has become ever more critical to narrow the interaction gap between them. An artificial agent capable of correctly inferring human intentions in real-time would not only allow the agent to improve its capacity to interact socially but would also allow the agent to adapt and respond to the user better as well.

As early as 1960, visionaries have repeatedly raised the notion of humans collaborating with artificial agents to form a 'collective intelligence'—with each counterpart enhancing the complementary strengths of the other [155]. Through significant breakthroughs in technology, we now live in an era where it is common for humans to interact with intelligent artificial agents. These interactions can range from making critical decisions alongside agents [79, 97, 203] to interacting with virtual agents through natural spoken dialogue [60]. However, current artificial agents are

still limited in their capacity to interact with humans effectively, especially when we start to consider the complexities present in everyday face-to-face human-human interaction. A primary issue lies in the fact that artificial agents are not yet capable of understanding the full range of human communication channels available, in particular, *human nonverbal communication* such as eye gaze, posture, gestures and facial expressions. In comparison, we humans can effectively interact with others to a better degree, primarily due to our inherent ability to simultaneously combine and process information from both verbal and nonverbal communication channels.

The introduction of gaze in multimodal human-agent collaboration presents an opportunity to improve an artificial agent's capacity to interact as the eyes play an essential role in the fluency of everyday human-human interaction. The human eye has evolved to be especially visible to provide nonverbal cues. Unlike any other species, human eyes have uniquely large, bright and clearly visible part of the eyes—the sclera [139]. This increased conspicuousness of the eyes affords an observer the ability to discern one's gaze direction easily [140, 239], allowing a collaborator to engage in behaviours such as *gaze following* (the ability and motivation to tell and follow where someone is looking by observing the direction of their eyes and estimating their line of regard [221]) or *joint attention* (the shared focus of two individuals on an object [177]). Further, as visual attention is intrinsically linked to action—meaning that humans look attentively at objects before acting on them [145]—an observer can make inferences about what actions their collaborator might perform next or simply gain insights into what they might be thinking. Likewise, an artificial agent, with access to the same nonverbal gaze cues, can potentially be trained to make the same inferences and predictions on the intentions of others.

In response to this potential, researchers have begun to explore the use of computational approaches (e.g. machine learning) to predict human intention through the use of eye-tracking, showing that gaze is an effective and promising predictor of human intention [26, 108]. However, existing implementations so far only have scratched the surface of what is possible. Studies so far have demonstrated the ability to *recognise* intentions from gaze detection, especially in the short-term (i.e. what the user will do next) and in specific tasks (e.g. sandwich making [15, 108]) but have not explored its use in interactive systems. Hence, when we start to consider the far-reaching benefits for the integration of socially interactive artificial agents in multiuser settings, we can begin to speculate how this might change the way we interact with agents that recognise

human intentions. Simultaneously, there is an ongoing body of research surrounding the use of gaze awareness in computer-mediated human-human interaction, showing clear benefits, especially in collaborative settings [12, 65]—including better coordination and communication.

This thesis seeks to push the possibilities of gaze input by combining the established benefits for incorporating gaze in collaborative settings while advancing the knowledge surrounding gaze-based intention recognition through computational approaches. The combination of these areas has direct benefits for its application in human-agent collaboration. At the same time, we can begin to address the limitations of current approaches explored in existing work, for instance, the issue of scalability when adopting a gaze visualisation approach (where it would be challenging to incorporate the gaze of more than two users without additional distraction). Therefore, the thesis not only aims to improve the capacity of agents to infer intentions by incorporating gaze input but also to improve the overall experience of human users when collaborating with agents that can recognise intentions. Hence, the thesis adopts a human-centred design approach—driven by a vision where humans and artificial agents can collaborate effectively in mixed-teams.

To scope the thesis, we have chosen a competitive gameplay setting, specifically using a contemporary multiplayer strategic board game called *Ticket to Ride*, for two primary reasons.

First, the chosen setting encompasses the crucial processes required for us to investigate the relationship between gaze and intention, and to collect a data set of natural gaze behaviours with ground truths in a controlled setting. A controlled setting allows us to control for the possible intentions in the space. In brief, players in *Ticket to Ride* are given goals at the start of the game and to achieve them, they will need to form hidden plans as an opponent can gain a significant advantage if they make correct inferences on them. However, as players make these plans in secret, their eyes give away this information, which we can capture using eye trackers. By analysing the data set on gaze behaviours in relation to their subsequent in-game actions, we can better understand the relationship between them for the purpose of intention recognition.

Second, the chosen setting allows us to switch up from single players competing against one another to collaborative teams competing against one another. This flexibility would then allow us to evaluate how mixed human-agent teams might benefit from the agent's increased capacity to collaborate with their human teammate. To do this, we must first provide the artificial agents with the ability to derive intentions from both the gaze their teammate and their competitors.

The overarching research question addressed in this thesis is:

> **Main RQ:** How might gaze-based intention recognition improve real-time human-agent collaboration in a competitive gameplay setting?

To answer the Main RQ, we structured the studies in this thesis into three parts. In the first part, we needed to understand gaze-based intention recognition better. We, therefore, conducted two human-human intention recognition studies to find out how humans display intentions while simultaneously evaluating the capacity of humans to infer intentions using gaze visualisation. In the second, we include the implementation details and evaluation of a gaze-enabled intention-aware artificial agent, developed as part of this thesis, that incorporates gaze into its intention recognition process. The evaluation, which uses the data set from the first part, demonstrates that incorporating gaze into the agent's planning process not only increases the agent's capacity to recognise intentions but also that it performs better overall than human observers. In the third and final part, we operationalise the agent by first giving the agent both the ability to communicate intentions of their opponent to its human collaborator and to explain its reasoning process if required. Subsequently, we evaluated the experience of the players playing the game with and without the assistance of the agent, which ultimately provided insights into how we can further improve the interaction between the human and an intention-aware artificial agent.

The findings of the studies resulted in three contributions towards the understanding of how artificial agents can support human-agent collaboration. In Chapters 5 and 6, the findings extend the relationship between gaze awareness and intention, by demonstrating that gaze when tracked over time, can lead to the detection of distal intentions (i.e. long-term intentions that often require several steps to be fulfilled [195]). Following, Chapters 6, 7 and 8 contributes to the design of a collaborative gaze-enabled intention-aware artificial agent and the demonstration of increased situation awareness through gaze awareness for human-agent collaboration.

The structure of this chapter is as follows. The two subsequent sections provides the background and motivations for the thesis (Section 1.1 and 1.2), which is followed by what we aim to achieve in this thesis (Section 1.3). Section 1.4 summarises the key contributions arising from the overall findings in this thesis. In the final section of this chapter, Section 1.5, we provide an overview of the subsequent chapters, including a brief description of how each informs the next.

1.1 Background

Since the mainstream adoption and proliferation of personal computers in the 1980s, humans have continued to harness its ever-increasing computing power. Today, over half the world's population depend on computing-based technology daily—to communicate, to learn, to work and to live in comfort.[1] Computers have evolved into far more complex machines than those initially adopted 40 years ago. Looking back further into the past before the PC revolution, we would find many future visions from modern computing pioneers that have yet to be fully realised.

One such vision is computer pioneer J.C.R. Licklider's *Man-Computer Symbiosis* [155] published in 1960. In this vision, computers would be able to work naturally with their human counterparts to solve problems that are not easy to address if attempted by either counterpart individually. Licklider explains that in the future, humans and computers would develop a 'symbiotic relationship', where the strengths of one counterpart would counterbalance the limitations of the other. To achieve this vision, Licklider has outlined several important prerequisites for the realisation of Man-Computer Symbiosis, which we have interpreted and broadly categorised in this thesis into *technology-centred prerequisites* and *human-centred prerequisites* (see Table 1.1 below).

Technology-Centred Prerequisites	***Human-Centred Prerequisites***
Speed mismatch between counterparts	Utilisation of nonverbal communication modalities
Memory hardware requirements	Matching the language between the counterparts
Memory organisation requirements	Ability to represent the mental state of the user
Input and output equipment	Complementary and effective division of labour

Table 1.1: Summary of prerequisites for the realisation of Licklider's *Man-Computer Symbiosis*.

Today, the technology-centred prerequisites outlined above are progressively being addressed as a consequence of rapid advancements in computer-based technology (in line with *Moore's Law*). While these technological advancements have enabled computers to become extremely powerful tools that are capable of performing a wide variety of tasks, the fundamental way we humans view and interact with computers and artificial agents by extension has generally remained the same throughout the decades—as mere *tools* to increase human productivity and efficiency.

In recent years, there is a current shift in thinking, which moves from the common view of computers as 'productivity tools' to *partners* (e.g [96]). This thinking resonates with Licklider's

[1] The recent COVID-19 pandemic has highlighted how much we humans depend on computing-based and digital technology in daily life, and at the same time has accelerated the need for improved human-centred technologies.

vision, where humans and computers should be partners to form a viable 'symbiotic partnership' that would ultimately result in a productive and thriving mutual partnership. At the same time, these technological advancements in areas such as *natural user interfaces* [86, 246, 255] presents new opportunities for changing the way we interact with computer-based technology. A timely example is the rapid development of intelligent voice user interfaces (VUIs) in recent years, which allow two-way natural interaction with devices using natural spoken dialogue [111, 199].

Although such developments are evidence of the desire to shift away from traditional inputs methods (e.g. keyboard and mouse) to more natural forms of inputs, the fact remains that we still interact with computers the same way—in an explicit *input-process-output pattern*. In contrast, the way in which humans interact when collaborating with others is inherently more dynamic, especially with the ability to gain mutual awareness of the situation and to establish common ground through negotiation or mediation [137]. With the ability to understand what others understand, humans can be proactive in their interaction in a collaborative setting, instead of merely waiting for instructions as with existing computers (master-slave relationship [95]). Such dynamic interactions are possible as humans not only have access to implicit nonverbal communication cues of their partners but also potentially possess a 'shared mental model' [121].

Reflecting on Licklider's *human-centred prerequisites* outlined in Table 1.1, we can begin to better understand the importance of each of the prerequisites for improving human interaction with AI, and how they are reflective of human-human interaction in real-world settings.

Among these prerequisites, the *utilisation of nonverbal communication modalities* prerequisite takes precedence in this thesis for three reasons. First, we posit that for artificial agents to work with their human collaborators effectively, artificial agents first need to harness nonverbal cues commonly present in human-human interaction. Second, addressing this prerequisite has immediate benefits for advancing multimodal human-agent interaction as augmenting current verbal-based communication technologies with implicit nonverbal communication will open up new avenues in the research field. Third, this thesis adopts the same shift in thinking where computers should be viewed and adopted as partners rather than tools.

To change the perspective of everyday human users, they first need to be shown that computers can become effective partners. Thus, enriching artificial agents with human-like cognitive capabilities given access to human nonverbal cues may be a viable approach for this demonstration,

especially within the context of human-agent collaboration where the humans can benefit from teaming with autonomous artificial agents.

1.2 Research Opportunity

Humans can easily recognise the potential intentions of others by simply observing their actions. For example, observing the action of a person picking up an empty coffee cup signals their likely intention to make a cup of coffee, or at the very least, the intention to make a drink. This example closely follows the current definition of intention recognition, i.e., the 'process of becoming aware of the intentions of other agents, inferring them through observed actions or effects on the environment' [102, 234]. Besides observing explicit actions that often occur after-the-fact, humans also inherently use a range of implicit nonverbal cues, including facial expressions, body language, gestures and eye gaze [138], to improve their ability to recognise the intentions of others. As such, observing a person gazing at the empty cup combined with their hand motion and gesture reaching towards the cup, would allow the observer to recognise the intention earlier.

A person's gaze alone, however, can give away much more information than just their next immediate action(s) (proximal intentions), especially when taking the broader environment into account. Building on the empty coffee cup example, not only can an observer make an inference on which empty coffee cup is to be chosen among others through their gaze, but also what the person plans to do with the cup after. It is possible for a human observer, prior to recognising the intention of picking up the cup, to make inferences on what drink the person intends to make (distal intentions). This prediction can be made by drawing comparisons on their observed gaze behaviour in the environment, such as whether the coffee machine or the tea cupboard have received more glances or which of the two have been looked at for a longer duration. As the eyes give away this rich information, they have on occasion been referred to as a 'window into the mind' [122, 221], as the direction of its gaze alone can convey several signals to an observer, including one's current point of attention and future intentions when observed over time.

A socially interactive artificial agent with access to the same explicit and implicit cues, combined with the ability to process them, has the same potential to improve its intention recognition capacity, especially when compared to approaches that use historical data for prediction. The introduction of gaze input for intention recognition in the context of human-agent collaboration

would allow the artificial agent to improve its collaborative capacity. The artificial agent can, for instance, provide proactive support knowing what the user intends to do in the long-term (e.g. providing alternatives), or even adapt its behaviour accordingly to help reach their common goals without the need for explicit instructions from the user. Moreover, it is possible for the artificial agent to be transparent about its reasoning process, simply based on its observations of the user's gaze, similar to how humans would do so using the same observed gaze behaviours.

In recent years, eye-tracking technology has begun to mature in terms of both its hardware and its underlying algorithms. These ongoing improvements, together with lower costs, have resulted in the technology increasingly being integrated into consumer devices such as laptops and head-mounted displays. Eye tracking capabilities can easily be added to existing devices using peripheral eye trackers, costing as little as €169[2] (at time of writing). In turn, this increased access to eye tracking technology has resulted in the increased exploration of gaze input in two rapidly growing fields—Human-Computer Interaction (HCI) and Machine Learning (ML). In HCI, eye tracking has been found to improve the communication and coordination between the users in computer-mediated human-human interaction by enabling gaze awareness. Gaze awareness, defined as *knowing what someone is looking at* [175], is typically produced by overlaying a live visualisation of the user(s) gaze over the interface (or environment), which is especially beneficial in remote settings where the eyes are absent [9, 63]. Whereas in ML, gaze has been shown to be an effective predictor of human intention [26, 108] as well as activity and context [229, 231].

Eye tracking has been shown to be beneficial in the advancing area of multimodal human-agent interaction, in particular in development and evaluation of early systems such as *iTourist* [201, 202]. These early experimental systems highlight that it is possible to produce interactive gaze-aware artificial agents can improve its capacity to support users with gaze input. However, they remain limited in their application as they are not yet fully capable of capturing the user's higher-level cognitive states, such as their *intentions*. Then again, as gaze input is an effective predictor of intention, it provides us with an opportunity to further expand beyond its use as an indicator of interest. Simultaneously, the investigation of gaze for intention recognition in multimodal human-agent interaction provides us with the opportunity to address the limitations found in previous computer-mediated human-human interaction studies. By selecting a collaborative

[2] https://gaming.tobii.com/tobii-eye-tracker-4c/

scenario where introducing gaze has shown to be beneficial, we can further begin to address Licklider's prerequisites as requirements for building an effective artificial agent partner.

To summarise our motivations, we determined that computers are not yet capable of understanding human behavioural cues despite the increased memory and processing speed to process the information, as they either do not capture them or do not use them well. Humans, on the other hand, are inherently better at interpreting nonverbal communication cues but are limited in their capability to process or store the large amounts of information potentially available to them. At the same time, the way humans interact with computers has essentially been unchanged over the past decades, where inputs are often provided to elicit a processed output. We argue that computers are often perceived as *tools* rather than partners due to the limited interaction capabilities between them, despite its prevalent use in modern societies in daily lives.

The introduction of eye tracking in human-agent collaboration setting presents an opportunity for us to bridge the gaps from both sides of the interaction as well as change the perception of computers as tools to partners. To elaborate, artificial agents can improve their collaborative capacity to provide better support by inferring the intentions of its human collaborator as well as of others (non-collaborators) through their gaze. In turn, the human collaborator can use the combined information that has been tailored based on the agent's inferences to make better decisions. Although eye tracking can provide novel ways of interacting with an artificial agent, we also need to factor in the considerations that may hinder the agent from becoming a trusted collaborative partner. Therefore, the thesis simultaneously addresses Licklider's *human-centred prerequisites* in order for the human-agent team to form a 'symbiotic relationship' [149, 155].

In future, human interaction with computer-based technology will only continue to become more ubiquitous as they continue to become smaller, inexpensive and available (in line with Weiser [251]'s *Ubiquitous Computing* vision). Thus, it is increasingly important to investigate how humans and computers can interact with each other, for instance, through implicit interaction [217]. One such paradigm that is closely relates this work (that encompasses implicit interaction) is Schmidt and Herrmann [218]'s *Intervention User Interfaces* interaction paradigm where computers act with autonomy and intervene when required. With input technologies such as eye tracking, we can start to explore what happens when we give computers the ability to observe their environment and situation implicitly and how they might *interact* in novel ways.

1.3 Aim and Scope

The aim of this thesis is to explore the potential for gaze to support collaboration between humans and artificial agents. To better scope the thesis, we have chosen to explore how gaze can support intention recognition, as knowing the intentions of a collaborator can improve the interaction between the counterparts, as shown in real-world human-human interaction.

As the existing literature surrounding human-agent collaboration has repeatedly argued that humans and computers can mutually benefit from cohesively working together, it is, therefore, crucial to focus our research efforts to narrow the existing interaction gap between them. Hence, the thesis focuses on extending the relationship between gaze and intention by providing the first steps towards building a human-centred artificial agent partner capable of increasing its capacity to recognise the intentions of its collaborator by incorporating their gaze into its planning.

The thesis simultaneously addresses all four of Licklider's *human-centred prerequisites* outlined in Table 1.1, with a focus on incorporating gaze as the primary form of nonverbal communication. By addressing these prerequisites, we can better design an effective collaborative artificial agent with minimal limitations, and in turn, allows us to demonstrate further the importance of incorporating nonverbal communication in the context of human-agent collaboration. Lastly, instead of exploring our aim using multiple collaborative scenarios, we have chosen to situate the studies in thesis using a single scenario: *collaboration in a competitive gameplay setting*. This choice, along with the use of a desktop-based environment instead of a physical co-located environment, is for several compelling reasons found in Chapter 3 (Research Design).

1.4 Thesis Contributions

The contributions arising from this thesis are threefold:

1. The thesis extends the relationship between gaze and intention, by contributing the knowledge that observing the gaze of human subjects over time allows humans and agents alike to recognise their plans (proximal and distal intentions). The thesis further demonstrates that gaze can be used deceptively when human subjects become aware that their gaze is being observed by others, particularly in adversarial settings. Section 5.10 provides a *taxonomy of gaze-based deception strategies* elicited from participants that attempted to obscure their

gaze to avoid their intentions from being known. In the future, it is possible that humans with the knowledge that AI systems are tracking their gaze would attempt to obscure their gaze. Hence, with the taxonomy, along with the data on gaze behaviours, we can begin to investigate how systems can disentangle between natural and deliberate gaze behaviours, and to design approaches and intelligent gaze-enabled systems that are robust to such behaviours.

2. The thesis presents the design considerations and initial evaluation of a working *collaborative gaze-enabled intention-aware artificial agent*—an artefact contribution. The agent is not only capable of inferring intentions using the combination of gaze and past actions more accurately and timely than human subjects but also able to explain how it came about its inferences using natural language afforded by a 'white-box' approach. Correspondingly, the thesis promotes the use of *human-centred design* approaches for designing and evaluating future multimodal explainable artificial agents. The development of the artificial agent and its capabilities in this thesis exemplifies how its design can be informed by human-centred studies in which we investigated how humans infer intentions and subsequently explain their reasoning process.

3. The thesis demonstrates the importance of incorporating gaze in the context of human-agent collaboration, afforded by a working prototype artificial agent with an improved capacity to infer human intentions in real-time through nonverbal communication. By incorporating gaze into the planning process of a 'white-box' artificial agent, it effectively extends the agent the interactive capabilities of the agent in three ways that are beneficial to human-agent collaboration. First, gaze awareness improves the agent's capacity to predict potential future states much earlier and more accurately (i.e. long-term/distal intentions). This allows a human in collaboration to adjust their plans accordingly to suit their goals, or confirm their own predictions, and therefore improve their ability to make better-informed decisions.

   Second, gaze awareness allows the agent to initiate the interaction by providing timely situation awareness updates (or queries) proactively. This consequently narrows the "interaction gap" between the two entities beyond the ability to recognise human nonverbal cues. This ability is especially beneficial when the agent is aware of the intentions of both the collaborator and of others, which would allow the agent to provide contextually-relevant information or recommendations that directly impact the goals of the human collaborator.

   Third, gaze awareness allows for the division of labour in human-agent collaboration. In

our work, we gave the delegated the task of intention recognition to the artificial agent to reduce the human collaborator's perceived cognitive load while maintaining their situation awareness. This division of labour, in turn, allows the human collaborator to better focus on their tasks and decision making (Chapter 8). Given the agent's potential to act autonomously, there are a multitude of ways the agent can provide assistance, for example, by taking over tasks of the user if the user is currently overloaded based on their gaze. However, we note that were able to demonstrate these capabilities as we factored Licklider's *human-centred prerequisites* as part of the considerations of the agent.

The discussion chapter (Section 9.3) provides a more detailed description of the contributions outlined in this section, along with its respective implications for future work and for the field.

1.5 Thesis Overview

The overarching research question addressed in this thesis is:

> **Main RQ:** How might gaze-based intention recognition improve real-time human-agent collaboration in a competitive gameplay setting?

This thesis is submitted as a *Thesis with Publication*[3]. Three published peer-reviewed papers are included. The first two papers are reformatted to the layout of the thesis for greater consistency, and form Chapters 4 and 5 respectively. The third paper contains two closely connected subsequent user studies and has been divided accordingly into two chapters (Chapters 7 and 8) for better clarity. Chapter 6 has been specifically written for the thesis but includes the contents of a published co-authored paper: [224] (prepared in collaboration with the primary author).

The thesis contains nine chapters in total; the introduction to the thesis, a literature review chapter, five study chapters, followed by a general discussion chapter. In each study chapter, we will introduce the study and how it related to the overall thesis and provide a summary of the findings in relation to the research questions formed in Chapter 2.

[3] https://gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

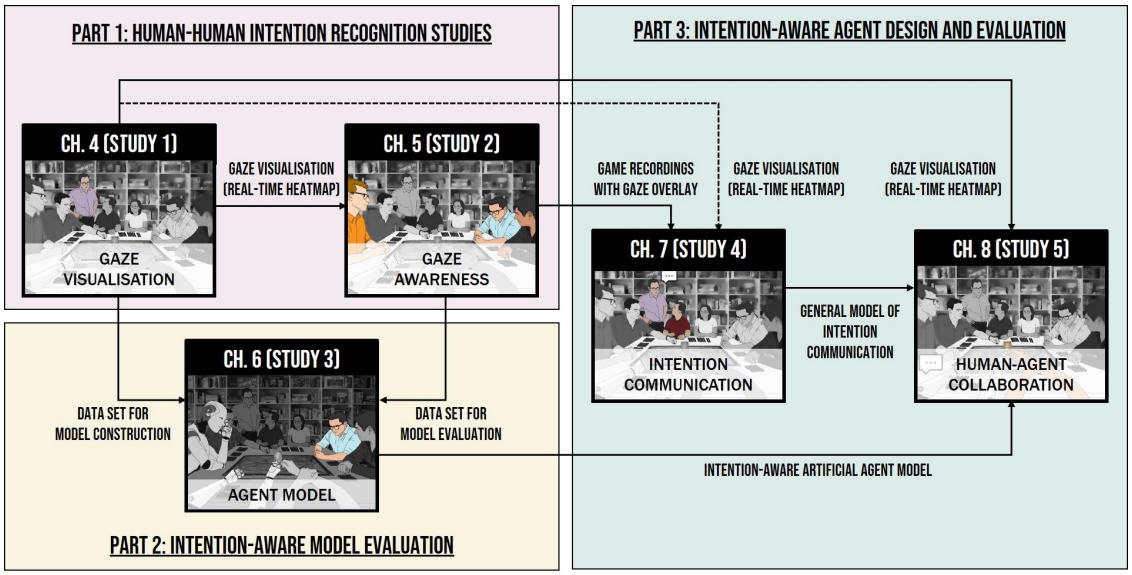


Figure 1.1: Thesis Research Structure.

Figure 1.1 illustrates the structure of studies in the thesis and how the findings of each inform the next. As depicted, the studies are structured into three main parts, each focusing a different set of actors from our imagined scenario (explained in Section 3.4). In Part 1, we conducted two human-human intention recognition studies (Study 1 and 2) to find out how humans portray intentions while evaluating the capacity of humans to infer intentions with gaze visualisation. These studies allowed us to gain a better understanding of gaze-based intention recognition while concurrently providing data sets for informing the agent model in the subsequent part.

In Part 2, we present and evaluate a proof-of-concept artificial agent prototype that incorporates gaze-based intention recognition into its planning process (Study 3). The results show that gaze not only increases the agent's capacity to intention recognition performance but also that the agent performs better than human observers in our previous human-human studies in Part 1.

In Part 3, we first operationalise the agent by giving the agent the ability to communicate intentions of their human opponent to their human counterpart and to explain its reasoning process if required (Study 4). We then evaluated the experience of the agent-assisted players playing an online real-time turn-based strategic game (i.e. *Ticket to Ride*) with and without the operationalised agent (Study 5). Both studies in this part focus on improving the collaborative capacity of the artificial agent through its design, given its ability to infer intentions, resulting in

findings that inform the next iteration of the artificial agent in which we leave for future work.

In the remainder of the section, we provide an overview of the chapters in the thesis, followed by a table that summarises the relationship between the study chapters and the research questions.

Chapter 2: Literature Review

Chapter 2 contains a literature review of five core research topics related to this thesis: (1) human-agent collaboration, (2) the role of gaze awareness in human collaboration, (3) the role of gaze awareness in HCI, (4) intention recognition, and (5) situation awareness.

The chapter is structured as follows. In Section 2.2, we present a brief overview of human-agent collaboration and its potential strengths and current limitations from a human-agent teaming interaction perspective. This section outlines the four human-centred prerequisites for achieving effective human-agent collaboration to inform the design of our proposed agent (Section 2.2.2–2.2.5). Section 2.3 highlights the importance of gaze in human collaboration, which reinforces the need to prioritise the *utilisation of nonverbal communication modalities* prerequisite.

Section 2.4 presents an overview of how researchers in human-computer interaction (HCI) have leveraged the increased availability of eye tracking technology. The section focuses on its collaborative use and includes the strengths and weaknesses concerning their applications and approaches. Section 2.5 presents an overview of how gaze has been investigated from an AI perspective with regards to intention recognition. In particular, how researchers in AI have approached the problem using computational approaches, which shows the potential for agent artificial to possess the ability to perform gaze-based intention recognition. Section 2.6 provides a summary of how human-agent teams can benefit from an agent that can provide situation awareness, especially from knowing the intentions of their human collaborators and of others.

Chapter 2 concludes by summarising the research gaps to be addressed in this thesis arising from the literature review, and the formation of the research questions in response.

Chapter 3: Research Design

Chapter 3 is divided into three subsections. In the first, we provide the justifications for the overall research approach chosen to answer the main RQ, including the justifications for selecting our

chosen scenario: *collaboration in a competitive gameplay setting* (Section 3.2). In the second, we present *Ticket to Ride*, a competitive turn-based strategy game employed throughout the studies in the thesis. This subsection includes the rules of the game, the concept of using eye tracking to improve intention recognition performance in the game and a feasibility study to determine its viability for the rest of the research (Section 3.3). In the third subsection, we first bring forward our imagined vision of the future where humans and agents can collaborate effectively in interactive social settings, in our case, around a table in teams strategically competing in a game of *Ticket to Ride* (Section 3.4). The section then presents the structure of the studies in this thesis, and how each informs the next in detail.

Chapter 4 and 5: Human-Human Intention Recognition Studies

Chapters 4 and 5 include the contents of two subsequent peer-reviewed papers, verbatim, but each has been reformatted to the layout of the thesis for consistency. The papers in their published format can be found in Appendix A. Chapter 4 (Study 1) has been published in the proceedings of the *2017 Annual Symposium on Computer-Human Interaction in Play* [188], while Chapter 5 (Study 2), which has been published in the proceedings of the *2018 ACM CHI Conference on Human Factors in Computing Systems* [184]. Both studies were designed as subsequent studies to evaluate the capacity of human subjects to perform intention recognition when given the ability to visualise the gaze of a tracked player, addressing the first RQ sub-research question below.

> **RQ1:** Does gaze awareness improve the capacity of a human player to perform intention recognition in a competitive gameplay settings?

Study 1 evaluates the characteristics of nine visual representations of gaze that would allow a human observer to improve their capacity to recognise intentions. The outcomes of this study reveal that gaze visualisations that balance real-time gaze information with a persistent summary of recent gaze behaviour, such as the *real-time heatmap visualisation*, was the most effective in revealing intent and was most highly preferred by the participants among the visualisations.

Study 2 focuses on the effect of gaze visualisation in competitive multiplayer gameplay. The objective findings show that players improved their capacity to uncover the plans of their opponent when shown a live visualisation of their opponent's gaze throughout the game. However, this

capacity was hampered when the opponent became aware that their gaze was being watched. The subjective findings indicate that humans have a limited capacity in performing gaze-based intention recognition, suggesting that the task may be more suitable for an artificial agent that is trained to process the rich multimodal information available in our setting. The knowledge and data sets arising from both chapters directly contribute to the next chapter, where we evaluate the performance of a prototype artificial agent that incorporates gaze into its planning process.

Chapter 6: Intention-Aware Agent Model Evaluation

This chapter presents an evaluation of an instantiated prototype planning-based artificial agent that combines gaze and explicit actions to perform online human intention recognition. The prototype agent model was constructed and evaluated using the accumulated knowledge and the data sets from the two previous chapters (Chapters 4 and 5), which evaluated *how* human subjects perform intention recognition when given the ability to visualise the gaze of a tracked player. The prototype agent model was constructed as part of an ongoing collaboration with the Artificial Intelligence and Autonomy Lab at the University of Melbourne, which aims at improving interactions between humans and agents through multimodal input[4]. Together, we have published the model, along with its formulation and initial results in the proceedings of the *2018 International Conference on Autonomous Agents and Multiagent Systems* [224].

Chapter 6 seeks to directly answer sub-research question RQ2 by evaluating whether a planning-based artificial agent model can improve its capacity to recognise the intentions of a human player by incorporating their recorded gaze information into its planning process.

> **RQ2:** Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?

The chapter finds that the prototype agent model, with the knowledge of where the human player is looking, can improve its intention recognition performance when compared with a based model-only approach in terms of making earlier predictions (inference horizon) and overall accuracy. Moreover, the findings show that the agent scored positively higher compared to a human interpreter using gaze visualisation, even when subjected to deceptive behaviours.

[4] https://cis.unimelb.edu.au/agentlab/human-agent-collaboration/

Chapter 7 and 8: Intention-Aware Agent Design and Evaluation

Both chapters represent Study 4 and Study 5, respectively. Both studies have been condensed and published as a single a peer-reviewed paper in the proceedings of the *2019 IFIP Conference on Human-Computer Interaction* (INTERACT 2019) [185]. The paper in its published format can be found in Appendix A. Both studies focus on enhancing the design of the prototype agent model presented and evaluated in the previous chapter, given its ability to improve its intention recognition capacity using gaze, responding to sub-research question RQ3.

> **RQ3:** How do we design gaze-based intention recognition systems for supporting human-agent collaboration?

The findings from both studies contribute to the understanding of how gaze can support human-agent teams in a dynamic collaboration scenario. Study 4 focuses on building the agent's capacity to communicate the inferred intentions, in our case, of a tracked human opponent to an assisted human player using natural language. Chapter 7 details a human-centred study where we employed an 'inverted' Wizard-of-Oz technique to elicit how humans convey and explain intentions when given the ability to perform intention recognition using gaze visualisation. The study resulted in a preliminary model of intention communication which informs how an artificial agent should communicate given the ability to recognise intentions. Study 5 operationalises the agent model in a competitive gameplay setting, which compares the experience of human collaborators playing an online strategy game with and without the assistance of the agent (Chapter 8). In this study, the agent adopts the preliminary language model of intention communication constructed in Study 4 to communicate with the human collaborator.

Chapter 9: General Discussion

Chapter 9 begins with a summary the overarching motivations of the thesis, the relationship between each study, and how they align with *human-centred prerequisites* outlined in Chapter 1. We then revisit and answer the research questions articulated in Chapter 2 (Section 9.2). The chapter then discusses the key contributions and implications of our findings arising from studies performed in this thesis (Section 9.3). This section is then followed by the limitations and future work (Section 9.4), and the concluding remarks (Section 9.5) of the thesis.

Thesis Chapters	Research Questions
Chapter 1: Introduction **Chapter 2: Literature Review** **Chapter 3: Research Design**	**Main RQ** How might gaze-based intention recognition support real-time human-agent collaboration in a competitive gameplay setting?
	RQ1 Does gaze awareness improve the capacity of a human player to perform intention recognition in a competitive gameplay setting?
Chapter 4: Study 1 Real-Time Gaze Visualisations for Human-Human Intention Recognition	**RQ1.1** How do we represent gaze to improve a human's capacity to perform intention recognition in a competitive gameplay setting?
Chapter 5: Study 2 Effect of Gaze Awareness on Human-Human Intention Recognition	**RQ1.2** What are the effects of gaze awareness for intention recognition in a competitive gameplay setting?
Chapter 6: Study 3 Intention-Aware Gaze-Enabled Artificial Agent Evaluation	**RQ2** Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?
	RQ3 How do we design gaze-based intention recognition systems for supporting human-agent collaboration?
Chapter 7: Study 4 Assistive Intention-Aware Artificial Agent Language Elicitation	**RQ3.1** How should an artificial agent communicate intentions given the ability to perform gaze-based intention recognition?
Chapter 8: Study 5 Assistive Intention-Aware Artificial Agent Real-Time Interaction Evaluation	**RQ3.2** Does delegating the task of intention recognition to the agent improve the player's decision making in a competitive gameplay setting?
Chapter 9: General Discussion	

Alone, we can do so little; together, we can do so much.

– Helen Keller

2

Literature Review

2.1 Introduction

This chapter aims to illustrate the importance of incorporating gaze in the context of human-agent collaboration by covering the five core areas of research: (1) human-agent collaboration, (2) the role of gaze awareness in human collaboration, (3) the role of gaze awareness in human-computer interaction, (4) intention recognition, and (5) situation awareness in human-agent teaming. The chapter is structured as follows. In Section 2.2, we present a brief overview of human-agent collaboration and its potential strengths and current limitations from a human-agent teaming and interaction perspective. This section outlines the four human-centred prerequisites for achieving effective human-agent collaboration to inform the design of our proposed agent (Section 2.2.2–2.2.5). Section 2.3 highlights the importance of gaze in human collaboration, which reinforces the need to prioritise the *utilisation of nonverbal communication modalities* prerequisite.

Section 2.4 presents an overview of how researchers in have leveraged the increasingly access to eye tracking technology. The section focuses on its collaborative use and includes the strengths and weaknesses concerning their applications and approaches. Section 2.5 presents an overview of how gaze has been investigated from an AI perspective with regards to intention recognition. In particular, how researchers in AI and machine learning have approached the problem using computational approaches, which shows the potential for agent artificial to possess the ability to perform gaze-based intention recognition. Section 2.6 provides a summary of how human-agent teams can benefit from an agent that can provide situation awareness, especially from knowing the intentions of their human collaborators and of others.

Chapter 2 concludes by summarising the research gaps to be addressed in this thesis arising from the literature review, and the formation of the research questions in response.

2.2 Human-Agent Collaboration

2.2.1 Overview

In this section, we provide a brief overview of human-agent collaboration and its potential strengths and current limitations from a human-agent teaming and interaction perspective.

The broader topic of 'Human-AI Interaction' has been much studied and discussed in the past two decades [29]. Here, we focus specifically on improving the collaborative capacity of artificial agents. Terveen defines collaboration as 'a process in which two or more agents work together to achieve shared goals' [237]. Bellamy et al. [29] explicitly point out that interaction is not the same as collaboration, as collaboration involves 'mutual goal understanding, preemptive task management and shared progress tracking', suggesting that collaboration has specific capabilities that both counterparts need to possess. Grosz [95] provides a comprehensive argument on the distinction between collaboration and interaction, and the importance of 'collaborative intentions' using a scenario with two mice in a maze. Grosz illustrates two mice trying to get out of the maze with one mice standing on top of the other, which shows that they are interacting but might not be necessarily collaborating or having the intentions to do so in the first place. The top mice using the bottom mouse as a step stool may benefit from this interaction by successfully escaping the maze but may not have the intentions to help the other mice to get out.

Extending Grosz's example, we can argue that the bottom mouse acted as a 'tool' for the top mouse to achieve its intended goals. Besides the fact the top mouse not having collaborative intentions, the bottom mouse did not express its intentions to collaborate, nor its intentions to get out as well. We can argue that there is a lack of communication between the two mice, and therefore did not communicate their intentions, shared goals or their capacity to collaborate. As such, we humans are often the 'top mouse' when we interact with computers and therefore perceive computers as 'tools' and not as collaborative partners, but at the same time computers often lack the capabilities to gain an awareness of its human user, especially in terms of their goals or intentions. Grosz, also, further makes a distinction between collaborators and tools, questioning whether we want computer systems to be mere tools, or 'something more' [95].

In recent years, there is a current shift in thinking, which moves from the common view of computers as productivity tools to *partners* (e.g [96]). The notion of a computer as a *partner* is an increasingly common metaphor for interaction design [29]. As partners, computers will help us get tasks done faster and perhaps also allow us to do more than we could alone. This notion of working together with computers as partners resonates with Licklider's vision of *Man-Computer Symbiosis* [155], where humans and computers should be partners to form a viable 'symbiotic partnership' that would ultimately result in a productive and thriving mutual partnership.

Previous research has shown many benefits for 'human–agent teaming', such as reducing operator's perceived workload, improving operator's task performance [56, 57] and maintenance of situation awareness [182]. A human-agent team is defined as '*a network where both intelligent systems (agents) and people cooperate effectively within one group, creating synergy by reinforcing each other's strong points and anticipate on each other's weak points*' [68]. This definition shares the same meaning with Licklider's *Man-Computer Symbiosis* vision, but use a more timely terminology instead of 'computers': *intelligent agents*. In the field of artificial intelligence, an intelligent agent is typically defined as "*an autonomous entity which observes and acts upon an environment and directs its activity towards achieving goals*" [212]. Among the many definitions of what an *agent* is (including intelligent agent, software agent), it is commonly accepted that an agent should possess the following characteristics [55]: (1) *autonomy*, (2) *observation of the environment* (through some forms of sensors), (3) *action upon an environment* (through some forms of actuators), and direction of its activity towards achieving certain goals [212].

In the previous chapter, we mentioned that despite the advances in computer-based technology and its capabilities to support human interaction with computers, the common view of computers as tools rather than partners have largely remained. We posit this perception is because computers do not yet possess the inherent capabilities that humans have to support collaboration, and therefore not often viewed as partners. To change this perception, intelligent computers need to possess not only autonomy but also characteristics that highlights its capacity to collaborate.

Nonverbal communication has been argued and shown to have benefits for human-agent teaming [5]. This is because nonverbal communication can reveal human mental states to an artificial agent such as knowledge, goals and intentions, and therefore allows the agent learn and perceive. As such, the primary aim of this thesis is to explore the potential for gaze—an important form of nonverbal communication—to support collaboration between humans and intelligent artificial agents. The awareness of a partner's gaze in human collaboration is an important nonverbal communication cue and signal which provides insights the mind of a collaborator, such as their current point of attention and their *intentions* (further elaborated in next section, Section 2.3). Therefore, by giving artificial agents the ability to not only perceive gaze but to understand its insights, it has the potential for changing the way humans interact and perceive computers.

Here, we can begin to ask, 'How do we go about giving artificial agents this ability?'. Grosz argues that researchers need to identify the capabilities that must be added to individual agents so that they can work with others and that these capabilities cannot just be 'patched on', but must be designed in from the start [95]. Hence, the thesis focuses on extending the relationship between gaze awareness and intention by providing the first steps towards building a human-centred artificial agent partner capable of increasing its capacity to recognise the intentions of its collaborator by incorporating their gaze into its planning. Teerveen [237] outlines two primary approaches for building the capabilities for human-agent collaboration: *human emulation*, which is more closely allied to AI and *human complementary*, which is closely allied to HCI. Following Teerveen, we argue for a unified approach where the agent can possess 'human-like' cognitive ability (emulation) but also complements the human collaborator with its potential to process intentions better than humans. However, there are a wide of considerations when designing for human-agent collaboration, which is highly dependent on the complexity, setting and goals. Barnes et al. [23] outlines five requirements for 'mixed-initiative' human-agent teams: a direct

link between the operator and a supervisory agent, interface transparency, appropriate trust, cognitive architectures to infer intent, and common language between humans and agents. On the other hand, Terveen [237] presents five fundamental issues that arise from collaboration, including agreeing on the shared goal(s) to be achieved, planning, allocation of responsibility, and coordination, shared context, communication, and adaptation and learning.

Returing back to Licklider's *Man-Computer Symbiosis*, which states computers would be able to work with humans to solve problems that are not easily addressed if attempted by either counterpart individually [155]. Licklider's vision not only presents a remarkably prescient prediction of the future, but the basic idea that humans and computer will communicate to collaborate and argues that computers should interact as well as communicate [149]. Moreover, Lesh et al. argues that most research in human-computer interaction does not seek to change the current style of interaction (i.e. *input-process-output*), but merely to make it more efficient through novels ways of interaction such as by using speech or direct manipulation [149].

Licklider [155]'s seminal paper outlines a number of important prerequisites for the realisation of the *Man-Computer Symbiosis* vision, which we have broadly categorised in this thesis into *technology-centred* prerequisites and *human-centred* prerequisites (summarised in Table 1.1). As mentioned in the previous chapter, the technology-centred prerequisites are progressively being addressed as a consequence of rapid advancements in computer-based technology. For example, researchers in human-computer have explored novel ways of interacting with computers through the developments of *input and output equipment.* However, the human-centred prerequisites have either received little attention or are usually explored in silos. In the following subsections (Section 2.2.2 – 2.2.5), we will describe the four human-centred prerequisites listed and argue the importance of addressing them together to enable effective human-agent collaboration.

2.2.2 Utilisation of Nonverbal Communication Modalities

> *Yet there is continuing interest in the idea of talking with computing machines. In large part, the interest stems from realization that one can hardly take a military commander or a corporation president away from his work to teach him to type. If computing machines are ever to be used directly by top-level decision makers, it may be worthwhile to provide communication via the most natural means, even at considerable cost.* (pg. 10, [155])

Our interpretation of the above excerpt from Licklider follows Lesh et al. [149]'s, where *communication via the most natural means* refers to the combination of verbal and nonverbal communication, which can be costly both in terms of computational power and communication overhead. In Lesh et al. [149]'s work, the authors explore a form of visible nonverbal communication called *gestures*, which is sometimes referred to as body language. Gestures are made with every part of the body (e.g. hands, face, posture of the body, body stance, legs, feet). The importance of gestures is well exemplified in human-computer interaction, especially when combined with other modalities (i.e. multimodality). An early example is Bolt [33]'s 'Put-That-There', which exemplifies the strengths and vast potential in the joint use of verbal (speech) and nonverbal (gesture) communication for offering an effortless way to interact with interactive systems.

Further, Lesh et al. brings up the use of *gaze* amongst other forms of nonverbal communication for achieving natural communication and collaboration, but specifically in its role for indicating engagement. Like gestures, gaze plays an important role in communication [5], such as revealing mental states, reinforce spoken communication, and provide additional information beyond what is being said (verbal communication) [93]. Humans can effectively interact with others to a better degree, primarily due to our inherent ability to simultaneously combine and process information from both verbal and nonverbal communication channels without significant effort.

2.2.3 Matching the Language Between the Counterparts

> *For the purposes of real-time cooperation between men and computers, it will be necessary, however, to make use of an additional and rather different principle of communication and control. The idea may be highlighted by comparing instructions ordinarily addressed to intelligent human beings with instructions ordinarily used with computers. The latter specify precisely the individual steps to take and the sequence in which to take them. The former present or imply something about incentive or motivation, and they supply a criterion by which the human executor of the instructions will know when he has accomplished his task.* (pg. 8, [155])

Communication is a key aspect in human-agent teaming [55]. One of the barriers to communication in human-agent teaming is the dissimilarities between their languages, especially for real-time cooperation, which Licklider [155] states "*may be the most serious obstacle for true symbiosis*". He explains that humans think more naturally and easily in terms of goals than

specific itineraries, implying the existence of human goals during communication. In short, instructions directed to computers specify courses (of actions) while instructions-directed to human beings specify goals. However, in our context, both the goals (what someone is trying to achieve) and 'courses' (steps to achieve goals) are important pieces of information and perhaps this can be both communication between the human and the agent in some manner.

However, enabling this symbiosis through mixed teams comes with significant challenges to effectiveness [137]. One such challenge is a lack of agent transparency, which hinders the human partner's ability to understand the decisions of the artificial agent [129, 178]. 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 [30, 55, 146, 249]. Therefore, to further improve human-agent collaboration, it is necessary for autonomous systems to infer human trust and workload and respond accordingly [8]. Further, researchers in AI argue that providing explanations supports transparency and may improve trust in the system [100, 161, 174, 192, 249]. Moreover, when using an agent as a decision aid, users would often seek an explanation of its output to improve their own decision making [247]. However, for an explanation to be effective, it must be at the right level of detail [131]. 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 [192]. Further, we need to consider the application domain, the audience of the explanation [98], as well as the *presentation format* (how to explain) and the *content* (what to explain) [75, 131].

2.2.4 Ability to Represent the Mental States of the User

Mental models are simplified representations used by individuals to explain and predict their surroundings [120, 211]. These models have content and structure or relationships between the content. Some examples of the content, which may be prior or gathered through perceptions (including inter-agent messages), are beliefs about the state of the environment and other agents, the agent's and/or others' goals, intentions, plans, knowledge, and reasoning rules that help an agent make decisions and handle events such as perceptions and communication.

During collaboration, the collaborator must have a mental opponent of their partner, as it allows them to adjust their level of collaboration. An issue lies in how the mental models are updated, in

most cases, this is achieved by simply communicating with one another but too much may lead to 'communication overhead'. Nonverbal communication has to reveal the mental states of others implicitly. For example, natural gaze behaviour can provide valuable information about user's cognitive processes and intention and its inclusion as an additional input modality, therefore provides great potential to improve human-computer interaction [21].

Byom and Mutlu [48] summarised three critical mechanisms for giving computer systems a *Theory of Mind*: (1) a shared world knowledge, (2) a mutual understanding of social cues, and (3) the ability of each to interpret the other's actions. While these mechanisms were studied independently using specific tests, it is believed that the combination of all three mechanisms is essential to creating an understanding between a human and an agent. In particular, the mutual understanding of cues and the ability to update their mental models are important factors for understanding the mind of another using the other two, allows for better collaboration.

2.2.5 Complementary and Effective Division of Labour

> *Computing machines can do readily, well, and rapidly many things that are difficult or impossible for man, and men can do readily and well, though not rapidly, many things that are difficult or impossible for computers. That suggests that a symbiotic cooperation, if successful in integrating the positive characteristics of men and computers, would be of great value. The differences in speed and in language, of course, pose difficulties that must be overcome.* ([155])

As humans, in general, can easily recognise where 'machines' such as computers would surpass them and vice versa [81], humans have always been quick to automate tasks where machines can perform better at, especially in terms of speed, accuracy and cost. This recognition has led categorisations of the general strengths and weaknesses of humans and machines, such as Fitt's HABA-MABA (humans-are-better-at/machines-are-better-at) approach, or simply, Fitt's list. For this reason, humans often view machines, however intelligent, as *tools*, as they merely perform tasks that are allocated to them. But what if computers are as effective at tasks that humans are good at, or even better, would this change their view? The problem lies in finding a task that both counterparts can mutually benefit, and intention recognition presents a good candidate. Further, we also need to manage the division of labour. One possible way is to allow *mixed-initiative* for flexible interaction, where each agent can contribute to the task what it does best [13].

2.3 The Role of Gaze Awareness in Human Collaboration

In the previous section, we touch upon the *utilisation of nonverbal communication modalities* prerequisite for achieving communication and collaboration with computers. Among the modalities, *eye gaze* presents an interesting collaborative modality for exploration, especially with its dual function in social settings—to signal and to perceive [91]. Eye gaze is one of the primary nonverbal communication signals, together with facial expression, posture, proximity, and tone of voice [62]. It is an extremely powerful signal as it alone can convey a number of signals to an observer, including interest, attention, emotions and intention. A person's gaze is one of the first cues humans look out for during face-to-face interaction, and where someone gazes upon is a typically an involuntary response to visual stimuli, which gives away a person's attention [262].

In this section, we focus on the importance of the role of *gaze awareness*—defined by Monk and Gale [175] as 'knowing what someone is looking at', in the context of human collaboration.

The appearance of the human eye gives humans the remarkable ability to identify what others are looking at, which subsequently allows them to engage in behaviours such as *gaze following*, that is, the ability and motivation to tell and follow where someone is looking by observing the direction of their eyes and estimating their line of regard [221], and *joint visual attention*—the act of looking where someone else is looking [47]. Unlike other primates, the human eyes have large, bright and clearly visible part of the eyes called the *sclera* [139]. In contrast, most primates have uniformly brown or dark-hued sclera, and this low contrast with the colour of their iris and facial skin makes it more difficult to discern the direction of their gaze from their eyes alone.

This difference in the appearance of the eyes has led Kobayashi and Kohshima to propose the *cooperative eye hypothesis*, which states our eyes have evolved to be especially visible to make it easier to coordinate when performing collaborative tasks [140]. This hypothesis was later examined by Tomasello et al. by comparing *gaze following* between various great apes and human infants [239]. The findings show that apes were more likely to follow the gaze of the experimenter using head movement. In contrast, infants followed the gaze more often when only the eyes moved. These findings suggest that humans depend more on eye movements than head movements when trying to follow the gaze of another. Infants develop the innate ability to engage in gaze following around 10-12 months of age [82], and within the first 18 months, they further develop the ability to engage in joint visual attention [47].

In face-to-face human interaction, gaze also serves as a deictic ("pointing") cue to an observer, which can be used to direct the attention of another person to an object implicitly or explicitly [221]. In essence, the eyes have two functions in social settings—to signal and to perceive [91]. Therefore, before the evolution of language, early humans may coordinate their activities and communicate their plans using their eyes along with other visually based gestures in accordance with the *cooperative eye hypothesis*. As a result, the use of gaze as a nonverbal communication tool becomes essential to our well-being, bonding and survival from an evolutionary perspective.

In collaborative settings, the knowledge of where someone is looking at (gaze awareness) has been demonstrated to be important for coordination with others, particularly in choreographing natural interaction (e.g. turn-taking in conversations [133]). The eyes and its movements are useful sources of information during social interaction due to its tight coupling with human cognitive processes [18, 83]. A collaborator can make inference based on their 'looking behaviour', especially when observed over time—the *eye-mind hypothesis* describes the relationship between where someone is looking and the amount of cognitive processing effort required [122]. This means that the longer a person gazes at an object, it signals that they are thinking about it (cognitive processing), and therefore represents their current point of attention or interest. As such, a collaborator can then adapt their behaviours to using this knowledge, which helps them to coordinate their actions better and to establish common ground if required [177].

Moreover, the ability to engage in gaze following and the use of gaze as a referential cue allows collaborators to engage in *joint visual attention* which is important in the development of *Theory of Mind* [101]. In brief, Theory of Mind is a pervasive human ability to attribute mental states such as beliefs, desires, goals, and intentions to oneself or another person [200]. This ability not only helps with the understanding of each other—an important factor in collaboration, but also allows humans can make predictions and reason about another person's behaviour [25, 200].

To sum up, gaze plays a significant role in human collaboration as well as in everyday social interaction. Gaze signals the basic direction of one's attention—often at things we desire in the environment [85]—and following the gaze of others over time enables the observer to discern their intentions [24], as humans naturally look at before acting on them [144]. As gaze allows a person to inferring the mental states of others, it allows them to form a 'theory of mind' about their collaborator's mental state and plans. As a result, the role of gaze has repercussion for

endowing artificial agents with a 'computational theory of mind' [215, 256], or at the very least for use to *represent the mental state of a tracked user*. In Section 2.5, we will further explore the concept of intention recognition, as gaze has been demonstrated to be an important cue for recognising intentions and for coordination and collaboration.

2.4 The Role of Gaze Awareness in HCI

In the previous section, we highlight the importance of the role of gaze awareness in human collaboration. In this section, we shift our focus to its role in human-computer interaction (HCI), focusing on two bodies of work that relates to this thesis. The first body of work is on computer-mediated human-human interaction, meaning that human user interacts directly with the gaze of others. At the same time, the computer only acts as a mediator between humans (*human-computer-human*). Primarily, the computer acts as a tool for creating gaze awareness, which is typically through gaze visualisation (Section 2.4.1). The second body of is on work where computers (or agents) has implicit awareness of the gaze of the user, and interpret the information to provide some level of direct assistance to the human user (Section 2.4.2). In addition, we include a brief review on gaze input in human-robot interaction (Section 2.4.3).

The rise in the adoption and exploration of eye tracking in HCI over the past decade is due to several reasons. First, eye tracking technology has begun to mature in terms of both its hardware and its underlying algorithms. These ongoing improvements, together with lower costs, have resulted in the technology increasingly being integrated into consumer devices such as laptops and head-mounted displays. Eye tracking capabilities can easily be added to existing devices using peripheral eye trackers, costing as little as €169[1] (approximately $200 USD at time of writing). Only until recently, eye trackers cost an upwards of $10,000 USD and were typically designed conducting lab-based studies to study human behaviour. Eye trackers provide computers with a user's gaze position in real-time by tracking the position of their pupils[2]. By processing the gaze positions, two basic spatio-temporal eye movements can be recognised: *fixations*, which are eye movements that stabilise the retina over a stationary object of interest and *saccades*, which are rapid eye movements used in repositioning the fovea to a new location in the visual environment [71].

[1] https://gaming.tobii.com/tobii-eye-tracker-4c/

[2] Duckowski [71] provides a comprehensive summary of eye tracking techniques.

2.4.1 Gaze Visualisation in Computer-Mediated Human-Human Interaction

Gaze visualisation is by far the most common approach for creating gaze awareness in computer-mediated human interaction. By overlaying a live gaze visualisation onto the interface (or onto the physical world [11, 104]) to create gaze awareness, it provides a complementary layer of nonverbal communication to an observer. As remote collaborations have been increasingly becoming common, it is only fair that many existing work have chosen to situate their work in this setting, especially given its unique challenges.

The benefits of gaze awareness through visualisation are also well pronounced in collaborative settings, where the eyes are not directly visible to their partner. This is because the presence of gaze provides an essential nonverbal cue for collaborators to monitor each other's attention and coordinate their actions as shown in traditional face-to-face collaborative settings [133], as described in the previous section. Hence, there is a substantial body of work surrounding the use of 'shared gaze' for collaboration in attempt to enhance the interaction. Gaze is typically shared by collecting the eye movement data from one person, and displaying over the interface of the other. In most cases, gaze is symmetrical shared, creating mutual gaze awareness (also called *shared gaze awareness*) as both users are mutually 'aware' where the other is looking. Gaze is typically used an indicator for a partner's point of attention, it often used as a dietetic referential pointer and coupled with another modality to improve coordination and communication. Hence, gaze is used to simulate the presence of cues in face-to-face interaction.

A popular remote collaboration task explored over the past decade or so is *pair programming* [63, 65, 119]. For example, Jermann and Nüssli has used the concept of shared gaze in a remote pair-programming scenario where both users can see the gaze of one another which allowed cross-referencing of sequences [119]. Other examples include collaborative search, for example, Brennan et al. have demonstrated that shared gaze can be more efficient than speech for the rapid communication of spatial information (spatial referencing) [37]. In general, the studies demonstrated that gaze sharing enabled pairs to refer to and find locations in shared code more effectively while strengthening shared awareness. More recently, eye tracking has been explored in the context of collaborative writing [143]. Kutt et al. showed that when incorporated into online collaborative writing, gaze sharing increased levels of collaboration, produced tighter work coupling and improved collaboration quality.

Gaze visualisation can further be augmented onto the real world for collaboration with physical tasks. For instance, Higuch et al. investigated how remote collaboration between a local worker and a remote collaborator will change if eye fixations of the collaborator are presented to the worker [104]. They tracked the collaborator's points of gaze on a monitor screen displaying a physical workspace and visualise them in the space by a projector or through an optical see-through head-mounted display. Among their findings, they found that eye fixations can serve as a fast and precise pointer to objects of the collaborator's interest and that the worker can predict some intentions of the collaborator such as his/her current interest. Similarly, Akkil et al. designed a system that projects the gaze of an instructor onto the workbench of a collaborator to perform an assembly task [11]. The study finds that collaboration became easier, referencing and identifying of objects became effortless, and that gaze awareness improved the partner's confidence that the task was completed accurately. Although gaze information flows one way in this instance, it showed the importance of augmenting real-time gaze information in the physical world. Furthermore, gaze visualisation has been explored in co-located collaborative settings, in particular, the work by Zhang et al. applied the concept of shared gaze awareness in a co-located collaboration scenario where gaze is presented on the same display to enhance collaboration between pairs [263]. They found that the knowledge of gaze to the pair eased communication, allowed them to be aware of their partner's interest and attention. They found that the awareness of gaze eases communication as pairs were aware of their partner's interest and attention.

The use of gaze in games have been a popular area for gaze interaction [243]. Hence, shared gaze through gaze visualisation has also shown benefit in both remote collaborative and competitive gameplay settings [147, 168, 205]. Špakov et al. [227] found that shared gaze increased mutual understanding between teammates and overall increased team performance. Maurer et al. explored how a shared gaze awareness changes how players work together and form collaborative strategies based on different combinations of gaze interaction and verbal communication [168]. Their findings highlight the positive influence of a shared gaze awareness on team cohesion and involvement towards increased perceived social presence among cooperating team members. Additional, gaze has also been explored in co-located gameplay settings—GazeArchers [197].

However, studies on shared gaze visualisations for various applications have found that the introduction of shared gaze awareness can be *disruptive*, often being referred to as 'distracting'

and 'confusing' (e.g. [66, 147]). We hypothesise that this is caused by two primary problems with gaze visualisation for digital interfaces. First, the way in which gaze has been visually represented in past works may be inadequate or ill-suited for the task and setting. This is expected, as the design of visual representations in graphical user interfaces is highly dependent on their context of use. This problem is exacerbated by an inherent problem with gaze: gaze is 'always on' [117], which makes designing gaze representations difficult as the technology cannot distinguish between focused scrutiny and more absent-minded looking.

Second, humans are not accustomed to interpreting visual representations of gaze, as the focal point of gaze is 'invisible' in normal interpersonal interaction. Gaze information can be 'noisy' due to the jittery movements and the way the eyes work (fixation-saccade-fixation cycle), and this is reflected in the representations. Gaze data is difficult to interpret and often misunderstood [66], as even trained professionals in eye-tracking evaluations make use of a range of gaze visualisations to make sense of the same collected data. Further, as gaze is fast-moving and never entirely still, any real-time visualisation overlaid on a user interface adds a highly distracting element that may interfere with the viewer's own thought process. As gaze visualisation is highly dependent on context and individual preference [65], software for visualising gaze in real-time often allows users to control its parameters such as by adjusting the colour, opacity and smoothness [38, 64] to mitigate potential negative effects and for added personalisation. For example, Tobii Ghost[3]—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.

In summary, the existing literature on shared gaze awareness makes clear that there is a need to balance *visibility*, *distraction* and *visual information*. Furthermore, we note that the problem with a majority of gaze representations is that they are designed to be used in post-processing of gaze data. Therefore, there is a need to design and evaluate novel gaze representations, such as that presented [153], to cater for settings where gaze input is not commonly used. Before we can design novel gaze representations, we need to understand the intrinsic characteristics of existing representations in relation to the context of use. Moreover, gaze in computer-mediated human interaction is often used as a dietetic referential pointer coupled with primary communication

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

modality such as speech, its presence on the interface alone can provide rich information from its observation. Knowing where someone is looking can provide rich information, from inference about their intentions to clues about their current cognitive activity based on their eye movement behaviours such as scanning, focused interest in an object, and repeated comparison of different objects [231]. In a recent study, gaze awareness has shown to improve collaboration efficiency in a collaborative assembly task, finding that the awareness of a partner's gaze significantly improved collaboration efficiency, as it gave them the ability to estimate their intent [248]. Hence, despite its numerous benefits for improving collaboration, we note that there are several issues that need to be addressed beforehand when adopting gaze visualisation in collaborative settings.

2.4.2 Gaze-Assisted Human-Computer Interaction

Gaze has shown to be beneficial in the advancing area of multimodal human-agent interaction, in particular in development and evaluation of early systems such as *iTourist* [201, 202]. These early experimental systems highlight that it is possible to produce interactive gaze-aware artificial agents can improve its capacity to support users with gaze. However, they remain limited in their application as they are not yet fully capable of capturing the user's higher-level cognitive states, such as their intentions. Then again, as gaze input is an effective predictor of intention, it provides us with an opportunity to further expand beyond its use as an indicator of interest. If computers become aware of our gaze, would they be able to provide assistance? Would they be able to address the limitations found in computer-mediated human-human interaction studies?

Qvarfordt, Beymer, and Zhai [201] and Qvarfordt and Zhai [202] applied gaze in a dialogue system, where a remote assistant is able to detect the interest of a remote user through their gaze patterns while conversing with them (remote user is unaware that their gaze is being tracked). Tracking gaze behaviour can serve as a natural deictic pointer and thus reveal interest in a natural way. The two-part work shows the potential for replacing human collaborator (or assistant) with an artificial agent. iTourist was developed as a stress test to investigate how much information can be gained from a user's eye gaze alone. One pattern to indicate interest is that a person looks long at one object with high intensity and long accumulated duration [202]. These studies also reveal other gaze patterns, such as when a user is no longer interested in a particular area.

In the games domain, implicit interaction using adaptive AI was found to be highly relevant to

our work, where the game AI would learn from the player's visual attention patterns to predict future actions. For example, Munoz et al. predicted player's actions in Super Mario Bros by analysing intentions from gaze data using an artificial neural network [180]. Similarly, Hillaire et al. used simple prediction models to determine the direction in which players were going to turn in a 3D environment based on gaze behaviour [105]. Likewise, gaming has been pushing AI applications such as by using probabilistic networks in a game to predict the player's next move to precompute graphics or by generating intelligent behaviours in non-player characters (NPC). For instance, Wetzel et al.'s AI game engine dynamically adapts its strategy (to maintain *flow*) based on the combination of eye movements and player's actions [253].

2.4.3 Gaze in Human-Robot Interaction

Eye gaze has shown to increase the efficiency and robustness in human-robot teamwork [169]. Gaze is an important social cue and signal in human-human interaction [16, 19], and this importance is reflected directly in human-robot interaction [42, 80]. Gaze in human-robot interaction offers two promising directions for enhancing interactions with robots and artificial agents. In one direction, robots can signal their own intentions by simulating head and eye movements [118, 176]. In the other direction, robots can monitor humans' eye movements to anticipate future actions [109] and to determine their level of engagement [4, 128, 136]. The use of gaze in the context of human-robot interaction has grown in the recent decade alongside improvements in eye tracking and computer vision technology. With exception of a few, the current body of work around gaze in HRI is typically aimed towards improving the collaboration between humans and robots by improving aspects of communication and coordination. These works generally fall into three broad areas: human subjects utilising the gaze of robots (e.g. [34, 80, 158, 181, 222]), mutual gaze between robots and human subjects (e.g. [6, 41, 259]), and robots utilising the gaze of human subjects (e.g. [7, 170, 206, 214, 242])—*our area of focus in this work.*

Intention recognition is an essential aspect in human-robot interaction [238], with applications ranging from collaborative assembly to assistive robots (e.g. [154]), however, this remains an open bi-directional challenge in the field [238]. This is well exemplified by Oliveira et al. [193]'s study with mixed groups of humans and robots. A critical aspect of gaze is its ability to signal intentions to others, whether intentionally or unintentionally ('*nonverbal leakage*' [181]).

As humans often look before performing an action, gaze provides a potential solution for intention recognition at two levels of complexity, the first being *engagement*. Robots can utilise the human's gaze to recognise the humans intent to interact with it. In Jayagopi et al. [118]'s work, the authors demonstrated the effect of human gaze and contextual dialog cues in helping a NAO robot determine whether a human was intending to speak to it and not to another person. They trained a Logistic Regression classifier to perform the addresee estimation and found that while gaze cues from the human speaker were the most important feature for addressee estimation, dialog-context features helps improve the estimation accuracy. This work highlights not only the strength of human gaze in intention recognition but also that a multimodal approach to intention recognition can be more effective.

The second level of complexity is for robots to fully utilise human gaze as a *predictor* of intention such as on objects that a human may be interested in. However, human intention is a complex construct, as it can be further classified into proximal (short-term) and distal (long-term) intentions [195]. Robots with the ability to recognise intention often lead to better performance when working on behalf of a human, as was demonstrated by Huang et al. [109]'s study. In their work, they trained a robot using Support Vector Machine (SVM) classifier to train a robot to recognise possible ingredients using gaze data they collected. In their experiment, a robot made smoothies for human customers and utilised the gaze of humans to predict the ingredients the customer was interested. The results showed that the robot picked the fruits that a human was considering quicker when the robot had access to gaze data; demonstrating the predictive ability of gaze.

2.5 Gaze-Based Intention Recognition

Based on the previous sections, we found that the concept of intention is underexplored in human-computer interaction [21]. Intention recognition is defined as the '*process of becoming aware of the intentions of other agents, inferring them through observed actions or effects on the environment*' [102, 234]. As visual behaviour is intrinsically linked to how humans plan and execute actions [144], researchers have leveraged the use of AI techniques to perform intention recognition from gaze, typically employing a machine learning approach (e.g. [26, 108]).

Researchers have begun to explore the use of supervised machine learning approaches to predict human intention through the use of eye-tracking, showing that gaze is an effective predictor of

human intention. As the thesis seeks to endow artificial agents with this ability, it is important to determine the selection of gaze features for building the models based on existing work [134].

Huang et al. investigated the predictive role of gaze in a collaborative task where participants role-played a sandwich-making scenario between a worker and a customer [108]. Using the ingredients eventually chosen as the ground truth for the customer's intentions, they measured the extent to which gaze cues served as predictors of their choices. They compared two approaches: (1) a simple attention-based intention predictor that performed predictions according to which ingredient the customer most recently fixated upon, and (2) an SVM-based classifier using four gaze features (number of glances, duration of first glance, total duration of glances, and whether a particular ingredient was most recently glanced at. The first approach outperformed random guesses achieving an estimated accuracy of 65%while the SVM classifier achieved an estimated real-time accuracy of 83%. Similarly, Bednarik et al. trained an SVM classifier using many gaze features to identify the type of task a user was performing when playing the 8-tiles puzzle game [26]. The types of tasks included *planning*, where the participant identifies the possible actions to take, and *cognition*, where the participant identifies the particular information the participant is currently processing. The gaze features employed were either *fixation-based* (e.g. fixation length, fixation count), *saccade-based* (e.g. count, area, duration), and *interface-based* (e.g. total no. of visited areas).

Other classes of algorithms have also been explored. For instance, Ishii et al. proposed an algorithm based on *decision trees* that estimated the user's conversational engagement with an artificial salesperson using attributes beyond gaze direction alone [114]. They trained a decision tree to serve as their engagement estimation model using four attributes—gaze direction transition, transition duration, amount of eye movement, and pupil size. Their model could predict a user's disengagement with an accuracy of around 70%. Similar to Huang et al. [108], Andris et al. also used a sandwich-making scenario but used a different approach to modelling prediction and intention, i.e. Epistemic Network Analysis (ENA) [15]. Their analysis gives an overall picture of the unfolding gaze patterns in dyadic collaborations.

Overall, such works, exemplify the use of computational techniques to harness the rich information available from the observation of gaze behaviour [26, 108]. However, these studies so far only have scratched the surface of what is possible, especially considering the far-reaching

benefits for its integration in socially interactive artificial agents in multiuser settings. Moreover, the works have demonstrated the predictions of short-term intentions (what someone might do next), and not in terms of intention recognition as plan recognition. Plan recognition is the term generally given to the process of inferring intentions from actions [51, 150]. Kautz and Allen define plan recognition as "the task of inferring intentions (in terms of plans) from the actions or utterances of observed actors" [130]. The recognition of plans have multiple benefits for human-agent collaboration, such as the reduction in user's communication burden [150]. We should also consider that intentions are multidimentional. An intention is '*a mental state that represents a commitment to carrying out future actions; involving mental activities such as planning and forethought* [36]. Pacherie summarises two common types of intentions found in the literature [195]: *future-directed intentions* and *present-directed intentions* in line with other researchers [36], *prior intentions* and *intentions-in-action* [219], *prospective intentions* and *immediate intentions* [35] and *distal intentions* and *proximal intentions* [172]. In summary, intention recognition is a complex process with many researchers presenting different models of intentions. In Chapter 6, we will elaborate how we approach the intention recognition problem.

2.6 Situation Awareness in Human-Agent Teaming

Situation awareness in human-agent teams have shown to be beneficial for human-agent collaboration [58]. Situation awareness is defined '*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*' [78]. Visual information helps people maintain up-to-date mental models or situational awareness of the state of the task and others' activities. This awareness can help them plan what to say or do next and to coordinate their utterances and actions with those of their partners [142]. Hence, we hypothesise that gaze awareness can lead to situation awareness in our setting, where players with knowledge of their opponent's plans (and how they might affect their own), can potentially improve the player's decision making. With the awareness of the current situation in terms of the past and present actions (i.e. what has happened so far) with clues of what might happen in the future, both the human and agent can adapt.

Chen et al. [58]'s Situation Awareness-Based Transparency Model (or SAT Model) acts as a model 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. According to the SAT model, each level describes the information an agent needs to convey in order to maintain a transparent interaction with the human. More importantly, an agent that can provide more accurate and timely projections (at the highest level) will help narrow down the potential predictions a player may possess. However, we note that this transparency comes at a cost. For example, Helldin et al. [103]'s study found that although transparency improved the operator's trust and performance, it caused decision latency and increased the operator's workload. Hence, it is important in our setting to balance the amount of information presented the player as well as how much assistance to provide.

2.7 Research Questions

The research topics covered in this chapter suggests that there are clear benefits for incorporating gaze into human-agent collaboration. The literature shows that gaze is an excellent indicator of intentions (Section 2.5). An artificial agent that can recognise intentions and subsequently reason and communicate them allows the human-agent team to gain a better awareness of the situation. Hence, there is a strong indication that intention recognition has the potential to support human-agent collaboration, especially if the agent can detect the intentions of others as well. Knowing the intentions of a collaborator and others (e.g. opponents) in the environment has multiple benefits such as the ability to update mental models, reduce the need to communicate, improve coordination and improve situation awareness for the human-agent team. Further, we situated the thesis in a competitive gameplay setting, as it presents an interesting controlled domain to explore (Section 2.4). As artificial do not yet possess the ability to perform gaze-based intention recognition, especially in our setting, this becomes the focus of in the first part of our research. Hence, the potential for gaze-based intention recognition to support human-agent collaboration forms the overarching research question addressed in this thesis is:

Main RQ: How might gaze-based intention recognition improve real-time human-agent collaboration in a competitive gameplay setting?

To help answer the Main RQ, we formed three sub-research questions, RQ1-RQ3. The first research question aims to drive the understanding of gaze-based intention recognition from the perspective from human users. Grosz argues that researchers need to identify the capabilities that must be added to individual agents so that they can work with others and that these capabilities cannot just be 'patched on', but must be designed in from the start [95]. Since human-agent collaboration is strongly inspired by how humans work together in teams, we wanted to find out how humans exhibit intentions and recognise them. To do this, we formed two further sub-research questions, RQ1.1 and R1.2. RQ1.1 address the gaze visualisation requirements to enable humans to recognise intentions, while RQ1.2 evaluates its effects in terms of increasing their capacity to infer intention recognition.

RQ1: Does gaze awareness improve the capacity of a human player to perform intention recognition in a competitive gameplay setting?
RQ1.1: How do we represent gaze to improve a human's capacity to perform intention recognition in a competitive gameplay setting?
RQ1.2: What are the effects of gaze awareness for intention recognition in a competitive gameplay setting?

In Section 2.4, we found that a major limitation in using human interpretation of gaze is *scalability*, where most works only focuses on two user scenarios, and we imagine it would be difficult to scale up to more than two users, which may add more confusion as users need to keep track who's gaze is which and where. Hence, we question what if an agent can perform this gaze-based intention recognition. Moreover, the current state-of-the-art employs gaze visualisation. While has shown some success in increasing task performance and improvements in coordination, we found that a number of issues. For instance, gaze becomes confusing when not in sync with other modalities such as speech, or can be distracting as it overlay over the screen at all time. Gaze has an inherent 'midas touch' problem and constantly moves as users on both ends scan the interface. Moreover, humans often find it difficult to interpret gaze visualisations, as they not accustomed to seeing the gaze of other humans in an explicit manner. This had led to studies investigating its appearance and also its use in different contexts. We determine that artificial agents can potentially assist a human with awareness of gaze. Therefore it is important to evaluate its capacity to infer intentions, which forms sub-research question RQ2.

RQ2: Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?

To answer RQ2, we will use collect the gaze behaviours as we evaluate the capacity of human subjects in RQ1. This approach also allows us to draw direct comparisons to determine if the artificial agent can outperform the human in terms of intention recognition performance. Therefore, for RQ1, we must consider how we can collect data with ground truths to answer RQ2. Moreover, RQ2 requires us to have a prototype of an artificial agent that combines gaze for intention recognition, and our review in Section 2.5 provides insights into how to begin.

The third sub-research question, RQ3, focuses on the collaborative aspects of the agent, in particular, its ability to communicate and its ability to provide situation awareness:

RQ3: How do we design gaze-based intention recognition systems for supporting human-agent collaboration?
RQ3.1: How should an artificial agent communicate intentions given the ability to perform gaze-based intention recognition?
RQ3.2: Does delegating the task of intention recognition to the agent improve the player's decision making in a competitive gameplay setting?

To answer RQ3, we formed two further sub-research questions, RQ3.1 and RQ3.2. RQ3.1 questions how an artificial agent should communicate intentions, given the ability to perform gaze-based intention recognition. As current artificial agents cannot yet infer intentions from gaze, this aspect is unexplored. We hypothesise that the agent will use different properties or even structures to communicate intentions, and therefore important that we investigate this aspect so that the agent can communicate effectively in a human-agent teaming scenario. RQ3.2 addresses the need to alleviate the user from the task of intention recognition. The addition of intentions of gaze-based intention recognition adds an additional layer of information the human player can draw on, but this can be detrimental as well, as outlined in Section 2.6. Hence, it is important to evaluate whether delegating the task to the agent can alleviate mental workload and improve decision making. RQ3 overall focuses on how we can design better artificial agents for human-agent collaboration, especially ones that integrate nonverbal communication such as gaze.

Future human-computer interactions will make it much easier for people and machines to work together. It will make it possible for machines to support people in cognitive tasks as well as physical tasks. The future holds a lot of possibilities for people and machines working together.

– Daniela Rus

3

Research Design

3.1 Introduction

In the previous chapter, we identified several issues pertaining to human-agent collaboration, along with its current limitations of current systems today. We argued that for artificial agents to collaborate effectively with their human counterparts, they first need to be able to utilise nonverbal communication cues during the interaction. The ability to track gaze unobtrusively using modern-day affordable eye trackers presents an opportunity to explore its use in the context of human-agent collaboration. In related work, shared gaze awareness has shown to be particularly beneficial in computer-mediated collaborative human-human settings, which leverages the importance of gaze awareness during humans interaction in the real world. Despite its many advantages for improving coordination and communication, many broader limitations remain unresolved, especially in terms of scalability and usability for future developments.

While existing work provides strong motivations for the incorporation of gaze into human-agent collaboration, they do not provide sufficient knowledge on how best to go about it. Notably, there is a lack of knowledge concerning gaze-based intention recognition, especially on enabling intelligent interactive systems to possess this capability. Hence, the thesis sets out to build this knowledge from the ground up, by first understanding how humans portray and infer intentions by giving them the ability to observe the gaze of others in our chosen controlled setting. We selected this as a first step towards building a gaze-enabled collaborative agent as being aware of the intentions of others is key for successful communication, as it not only reduces communication overhead but improves the overall collaboration experience.

This chapter is divided into three subsections. In the first, we present our overall research approach to the problem, along with the justifications for selecting a scenario-based design approach and our chosen scenario: *collaboration in a competitive gameplay setting* (Section 3.2). In the second, we describe *Ticket to Ride*, a competitive turn-based strategy game that we employed throughout the studies in the thesis, including the rules of the game, the concept of introducing eye tracking in the game to investigate gaze-based intentions and a feasibility study to determine its viability for our research (Section 3.3). In the third, we outline our imagined vision of the future where humans and autonomous embodied agents can collaborate effectively in an interactive social setting, in our case, around a table in teams strategically competing in a game of *Ticket to Ride*. As we employed a scenario-based approach, we can form multiple studies focusing on different actors and activities stemming from our imagined vision. Hence, in Section 3.4, we provide a detailed structure of the studies included in this thesis, including how the findings of each motivate and informs the next, and how they fit into the aims of the thesis.

3.2 Research Approach: Scenario-Based Design

The overarching research question addressed in this thesis is:

> **Main RQ:** How might gaze-based intention recognition improve real-time human-agent collaboration in a competitive setting?

To answer this question, we adopted a human-centred design approach, in which we focus on the behaviours and experience of humans rather than their collective needs and requirements. This

shift in focus is due to the thesis aims, where we set out to explore how incorporating gaze-based intention recognition in the context of human-agent collaboration can potentially narrow the interaction gap between humans and artificial agents. To do this, we must first address the fact that socially interactive agents do not yet possess the capability to perform gaze-based intention recognition, which requires us to study how humans perform this inherent ability in the first place. Hence, it becomes less of a priority in this thesis for users to participate in shaping the design of the artificial agent in the earlier stages. Our overall human-centred approach is, therefore, more focused on *human-emulation* [237] to endow the artificial agent with human-like abilities. Once we address this unique challenge, we can only then start to investigate how an artificial agent given both the ability to recognise and communicate intentions can support a human player collaboratively (as performed in Chapter 8). Meanwhile, in Chapters 4-7, we focus on eliciting natural human behaviours to inform, design and evaluate the artificial agent.

As our overarching human-centred design approach, we employed the use of *Scenario-Based Design* to design a series of studies to inform the different capabilities of a collaborative artificial agent. Carroll describes Scenario-Based Design (SBD) as a "a family of techniques in which the use of a future system is concretely described at an early point in the development process" [53]. Scenarios, narrated as simple stories, help us to communicate problems and ideas, can help in envision outcomes, and provide a basis for testing and stimulate thinking about possible outcomes. Further, scenarios can be abstracted, categorised and can be written at multiple levels of granularity [52]. As a consequence, the open-ended and informal narratives of what people do and experience as they try to perform different activities with or without making use of a computer application has made it a popular approach for designing user-centred systems [126].

Our adoption of scenarios for building the capabilities of an 'intention recognition system' is akin to the early work on speech recognition systems, where scenarios were adopted as part of its design process and for eliciting patterns (e.g. [125, 126, 127]). The prevalence and advancements in intelligent voice user interfaces would not otherwise be possible without the successful outcomes from such studies. With this in mind, we further observe that our approach mirrors the way voice interfaces have developed over the decades. Giving computers the ability to 'listen' (through microphones) or to 'speak' (through speakers) mirrors the ability to 'see' the gaze of the human through eye trackers. Over time, the ability to interact using voice has allowed us to

study the patterns of speech interaction, leading to significant developments in its recognition and simulation. Henceforth, we anticipate that the adoption of a scenario-based design approach would allow us to follow a similar trajectory, but with eye tracking and intention recognition.

Following Wolf [257], we focus on the four main elements when producing a scenario: (1) actors, (2) background information on the actors and assumptions about their environment, (3) actor's goals or objectives, and (4) sequences of actions and events. Typically, the SBD approach is used to describe a human actor's interaction with a system. In this work, we extend actors to include artificial agents, as we view them as having the same cognitive abilities as the human actors.

Figure 3.1: Envisioned Human-Agent Collaboration Scenario – *Collaboration in a Competitive Setting.*

Figure 3.1 illustrates our chosen overarching scenario, a vision of the future in which the thesis is working towards to, where humans and autonomous embodied agents can interact with each other naturally in a playful competitive setting. The figure shows a series of possible interactions between the actors, which is investigated in this thesis. Most importantly, the figure shows the perspective of collaboration between an artificial agent (represented as a humanoid robot) and a human player collaborating in the game against other human-agent teams. Here, the human

player and agent are in the process of strategic planning, but also simultaneously inferring their opponent's strategies observing both their explicit actions and the implicit nonverbal signals to infer their potential strategies to inform their planning. One of the many strengths of adopting an SBD approach is its flexibility to describe scenarios at different levels of granularity. Hence, it allows us to focus on different interactions between the actors that may occur in the context scenario, along with their actions and tasks. This will be further elaborated later in Section 3.4.

The envisioned overarching scenario takes on a speculative narrative that, in turn, helps us to form 'what-if' scenarios, along with follow-up questions that we can use as a basis for further investigation. Such questions can include: *What if artificial agents can infer the intentions of human players from observing their gaze? Would they be able to explain how they came about their inferences by describing what they observed?* Questions from a human perspective include: *Would humans change the way they look knowing an artificial agent constantly watches them? How would we go about building artificial agents that can effectively collaborate alongside humans?*

Further, by envisioning an artificial agent in our setting, it highlights its current limitations from an interaction capability perspective. For example, the agent cannot yet infer the nonverbal communication cues of its teammate and its opponents but given the ability to do so would allow the agent to adapt its behaviour or plans accordingly, to provide proactive support to the human counterparts for the team to reach their common goal [236], or even to be transparent about its reasoning process based on its observations. At the same time, it also shows that the agent cannot yet communicate its inferences, as it lacks the protocols to do so. On the other hand, the presence of human players in the scenario helps keep the humans in sight, enforcing us to always to consider the human players when designing the capabilities of the artificial agent.

One of the factors that led us to adopt an SBD approach is that it is reflective of current works in human-computer interaction that adopts a use case application to demonstrate the ability of the system. One such example that leveraged eye tracking is the development *iTourist*/*RealTourist* [201, 202]. In this two-part work, the authors use a tourist planning scenario together alongside a human-centred approach to elicit both human gaze and conversational patterns to inform the design of a system that can engage in conversation based on the user's gaze patterns. Using such approach as a basis, therefore, allows us to expand the current application of eye tracking further, moving away from gaze as a referential cue to signal interest as used in works that employ gaze

visualisation, and towards the ability to make inferences on intentions.

The primary reason for selecting a competitive multiplayer gameplay setting is that we required a controlled setting where we can isolate gaze behaviours from other forms of communication (e.g. verbal communication). In competitive games, players can play out the entire game without explicitly communicating with one another. From a data perspective, it allows us to build a record of natural gaze behaviours, changing states of the game, and most importantly, ground truths through their explicit actions. From a computing perspective, it allows us to easily abstract the rules of the game into the agent model, allowing us to evaluate different scenarios with the data and further explore the interaction capabilities of the agent.

In the next section, we describe the strategic board game illustration in Figure 3.1—*Ticket to Ride*, which we have selected from a list of candidate games. Competitive strategic multiplayer games and gameplay have long been used to evaluate the performance of artificial agents [261], ranging from classic board games (e.g. Chess [50]) to multiplayer real-time strategy games (e.g. StarCraft [194]). A popular example is AlphaGo developed in 2015 by DeepMind Technologies to play the two-player turn-based strategy board game *Go*, one of the world's most complex games. In 2016, AlphaGo defeated Lee Sedol, one of the world's best Go players and has since undergone multiple successions [223], and approach being applied to other domains (e.g. [220]). Two decades earlier, IBM's supercomputer Deep Blue defeated grandmaster Garry Kasparov, the greatest chess player at that time [50]. Kasparov defeat resulted in *centaur chess* (or cyborg chess), where players on both sides collaborate with computers to increase the level of play to new heights [27].

Unlike Chess and Go, *Ticket to Ride* is a game of imperfect information where players hold cards containing their hidden goals and resources and therefore do not know the full state of the game [213]. A good example of imperfect information games is a card game where the player's cards are hidden from the rest of the players. Imperfect information makes things much harder for the AI in terms of making predictions but better reflective of real-world applications [141]. A recent example of AI that is tasked with imperfect information is *Libratus* [40], a computer program designed to play poker, specifically heads up no-limit Texas hold 'em. Similar to our goals, Libratus is intended to be generalisable to other, non-Poker-specific applications. As such, Libratus is only able to learn any game where opponents may be hiding information or even engaging in deception, but also in any situation in which incomplete information is available.

3.3 Ticket to Ride

Ticket to Ride (USA) is a popular 2-5 player turn-based strategy board game designed by Alan R. Moon and published by Days of Wonder in 2004. Ticket to Ride has won multiple international awards, sold over 6 million copies and translated into 18 languages. In this game, players collect cards of various types of train cars that enable them to claim railway routes connecting cities across a map of North America. The objective of the game is to accumulate the most amount points throughout the game, as the player with the highest number of total points, calculated at the end of the game, wins the game.

We chose this game for several reasons. First, it is a widely-played board game, already familiar to many, with a digital game version that is almost identical to the board game[1]. It has relatively simple rules involving clear goals and actions, which can be learned in under 15 minutes by interested players who have not played it previously. This simplicity is important as games with overly complicated rules may confuse players and unable to provide the data we need for evaluating intention recognition. The game also has a relatively short gameplay time, approximately 30 minutes for a two-player game and shorter for experienced players.

Second, Lankes et al. have suggested that in competitive settings, turn-based games are better-suited for overlaying the gaze of an opponent on the interface, as opposed to real-time first-person shooters [147]. This suggestion is because such games allow players to alternate between playing their own moves and watching their opponents' behaviour as they take their turn. Third, *Ticket to Ride* is a map-based game, in which the players' focus and strategic calculations are closely concentrated on the geography of the map; gaze awareness and eye movement studies with map-based applications have been shown to be highly successful, as the static map gives users a point of reference [135, 202, 263]. Fourth, it is a game of imperfect information, and therefore approaches arising from this work may potentially be generalised to other real-world scenarios, as seen with the progression of *Libratus* using poker as a form of imperfect information [40].

Lastly, *Ticket to Ride* has been a popular game in the AI community from different perspectives, from its use for educating theories [157], to its use as an example of playtesting [173], and for running simulations on different agent models [112].

[1] https://www.daysofwonder.com/tickettoride/en/

3.3.1 Gameplay and Rules

At the beginning of the game, each player is randomly dealt three *destination ticket cards*, which are unknown to their opponents. Each destination ticket card shows a pair of cities (e.g. from Dallas to New York) along with a point value in which the players will attempt to complete in secret, forming their hidden goals. The player must keep at least two of ticket cards and discard the unwanted ticket card to the bottom of the stack, if any. The ticket cards that are kept may not be discarded for the rest of the game. There are 30 destination ticket cards in total and are a mix of short, medium and long connections reflected by their point value, which ranges between 4 points (Denver to El Paso) to 22 points (Seattle to New York).

Players are also randomly dealt four *train car cards* from a random pool at the beginning of the game as their starting hand. These cards act as resources for claiming routes on the board and can be one of nine different colours: black, blue, green, orange, purple, red, yellow, white or wild. A wild card, also known as a locomotive card, is a special type of card that can be used as if it were of any other colour. Each colour (except for wild) directly corresponds to the colours of the routes located on the map, as shown in Figure 3.2. Each player also selects a colour which corresponds to a group of 45 coloured train pieces with a matching scoring marker.

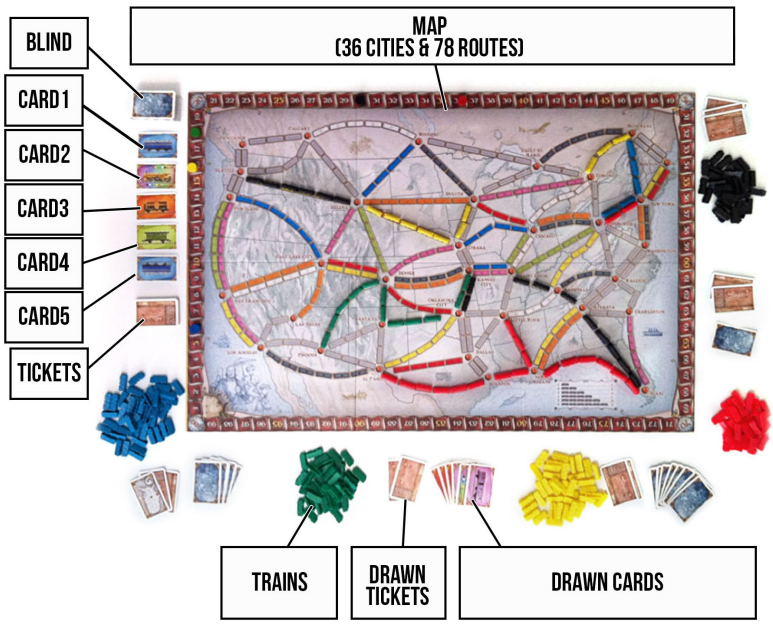


Figure 3.2: Ticket to Ride Board Game (and its areas of interest and components).

The objective of the game is to score the highest number of total points calculated at the end of the game. The player with the most points wins the game. Players score points in three ways:

1. **Claiming a Route**. Points are awarded when a player claims a route between two adjacent cities on the map. The number of points depends on the length of the route in terms of train cars, and are as follows: 1 = 1 point, 2 = 2 pts, 3 = 4 pts, 4 = 7 pts, 5 = 10 pts and 6 = 15 pts.
2. **Completing a Destination Ticket Card**. Players obtain the corresponding points stated on the destination ticket cards if they successfully connect them in a continuous path and will lose the same number of points if they do not. Players can draw additional ticket cards during their turn in the game if desired.
3. **Completing the Longest Continuous Path of Routes**. The player with the longest continuous path of routes at the end of the game is awarded an additional ten bonus points.

On each turn, the player must perform one (and only one) of the following three actions:

1. **Draw Train Car Cards**. The player may draw two train car cards. The player may take any one of the five face-up cards (Card 1- Card 5 in Figure 3.2) or may draw the top card from the deck, which is a blind draw. When a player draws a face-up card, the card is immediately replaced with a card from the deck. If the player draws a face-up wild card from one of the five face-up cards, the player may only draw one card instead of two. However, if a player is lucky enough to get a wild from the top of the deck in a blind draw, it stills counts as a single card and the player may still draw a total of two cards that turn. For the second card, the player can draw either from the face-up cards as long it is not a wild card, or from the top of the deck. In short, players can either draw two train cards of any colour combination, or one wild card that is faced-up during their turn.
2. **Claim a Route**. The player may claim a single open route between two cities on the board by using up their train car cards, and is never required to connect to any of his previously played routes. To claim a desired route, the player has to discard the number of cards of the same colour from their hand. The colour and quantity of cards are determined by the desired route they are trying to claim. For example, the route between Denver and Helena requires four green train car cards to claim, and therefore requires the player to have four green train car cards (or a combination with wild cards) in hand when claiming, which will be subsequently be discarded. After discarding the cards, players place their coloured train

pieces on the route to mark as theirs. Out of the seventy-eight routes, twenty-two routes are grey coloured, which indicates that the route can be claimed using a set of cards of any one colour. A route claimed by one player can no longer be claimed by any other.

3. **Draw Destination Tickets**. Players have the option to draw additional destination tickets cards during their turn. To do this, the player draws three destination ticket cards from the top of the respective deck. The player must keep at least one of them, but he may keep two or all three if desired. Any returned cards are placed on the bottom of the deck.

The game reaches an end when one player is left with two or fewer trains pieces at the end of a turn, at which point all players proceed to take one more turn. The game then ends, and all players calculate their scores. The player with the highest score wins the game.

In summary, *Ticket to Ride* is a 2-5 player turn-based strategy board game where players compete to build train routes between adjacent cities across a map of North America. Players gain points for building a connection between specific pairs of cities that they are randomly assigned on destination ticket cards, which are unknown to the opponent. Players lose points if they fail to complete ticket cards they hold, and can block each other's paths, as only one player can claim each train route. 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. Keeping information hidden is, therefore, core to the game, as a player can gain a significant advantage by correctly guessing their opponents' hidden objectives.

3.3.2 Concept

Figure 3.3 illustrates a simple example of how knowing the gaze of an opponent can improve a player's capacity to infer intentions in the game of Ticket to Ride from a first-person perspective. Hence, imagine that your opponent is attempting to connect a route between Santa Fe and one of the other cities on the map in the figure. The intention recognition problem is to determine the destination city (potential goal) as well as the routes the opponent might take to reach the destination city (plans). On the left figure (without gaze), you can see that the opponent has already traversed (or claimed) the route from Santa Fe to Denver. Thinking rationally, we can argue that it implies that the probability of the opponent's final destination being Oklahoma City is smaller than that of any other city. If the opponent wanted to connect to Oklahoma City, then

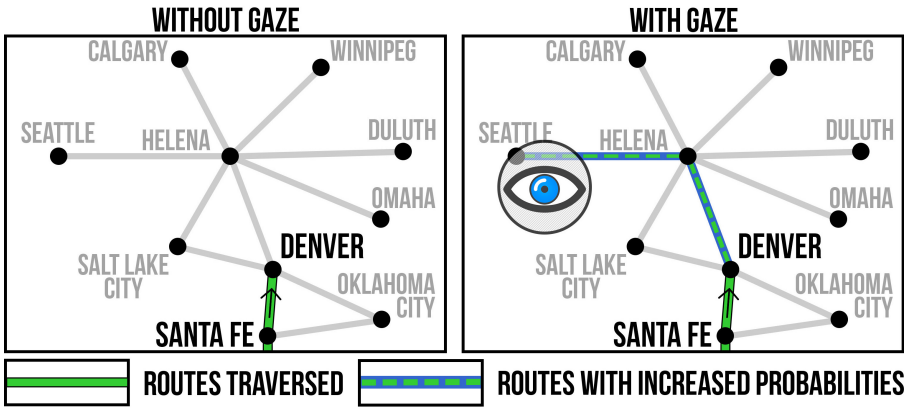


Figure 3.3: Gaze-Based Intention Recognition Concept (in relation to *Ticket to Ride*).

they would have traversed from Santa Fe directly to Oklahoma City instead of going through to Denver unless it is part of another plan to head in that direction.

From this single claim route shown, it is not very easy to distinguish the probability of the distant cities in the north (Seattle, Calgary, Winnipeg, etc.)—they are all the same distance from Denver (i.e. two moves away). The example on the right (with gaze) shows how gaze might increase the intention recognition capacity in this game. Here, we can see that the opponent is gazing upon both the Seattle and the Seattle to Helena route. From this observation, we can now argue that this represents a potential future action and that Seattle is a more likely the final destination compared to Calgary or Winnipeg, especially if these cities have received no interest through their gaze. With the knowledge, we can determine that the opponent is likely to claim the Denver to Helena route next. We can also argue that Calgary is still more likely than Oklahoma City, which neither fits our observed in-game actions nor our observed gaze actions.

3.3.3 Feasibility Study

Following through with the concept presented in the previous section, we conducted a feasibility study to determine the suitability of *Ticket to Ride* for eliciting gaze behaviours for the purpose of intention recognition further. To do this, we recorded a playthrough of the game using the physical board game version with four players. Initially, only one player knew the rules of the game, hence a quick test round was played so that all players became familiar with the rules and procedures in the game. The subsequent round is a real match which lasted one hour

Figure 3.4: Feasibility Study Experimental Setup.

(approximately). One of the players wore a pair of Tobii Pro Glasses 2[2] eye tracking glasses during the study. Figure 3.4 shows the study setting with players seated around a table, recorded from four different camera angles. From these five camera feeds (4 fixed cameras + 1 wearable tracking), we can observe the actions and behaviours of each player. We reviewed the recordings from both the camera feeds and the eye tracking glasses. From the recordings, we observe three key behaviours:

Behaviour 1: Hiding Intentions

The player will determine the best possible next steps without taking too long to not give away his intentions. This shows that players can become aware that other players are watching their actions or nonverbal cues such as their gaze in this setting as well. In general, it is also possible that players might not want to keep other players waiting, especially if it is during their turn. Hence, players would often have a plan in mind but do their best to keep them hidden as opponents can gain a significant advantage if they correctly guess their plans.

[2]https://www.tobiipro.com/product-listing/tobii-pro-glasses-2/

Behaviour 2: Early Planning

Figure 3.5 shows that the gaze of a player can reveal the plans of a player early in the game. As shown, the player forms his initial plans at the start of the game when they receive his tickets, forming his goals. As the game progresses, the player repeatedly checks his plans, often looking back and forth between the cards in his hand and the area on the board to determine if he has enough resources to complete that part of the plan. When the player's gaze is combined with his claimed routes, it can be used to reveal the trajectory of his intended plans.

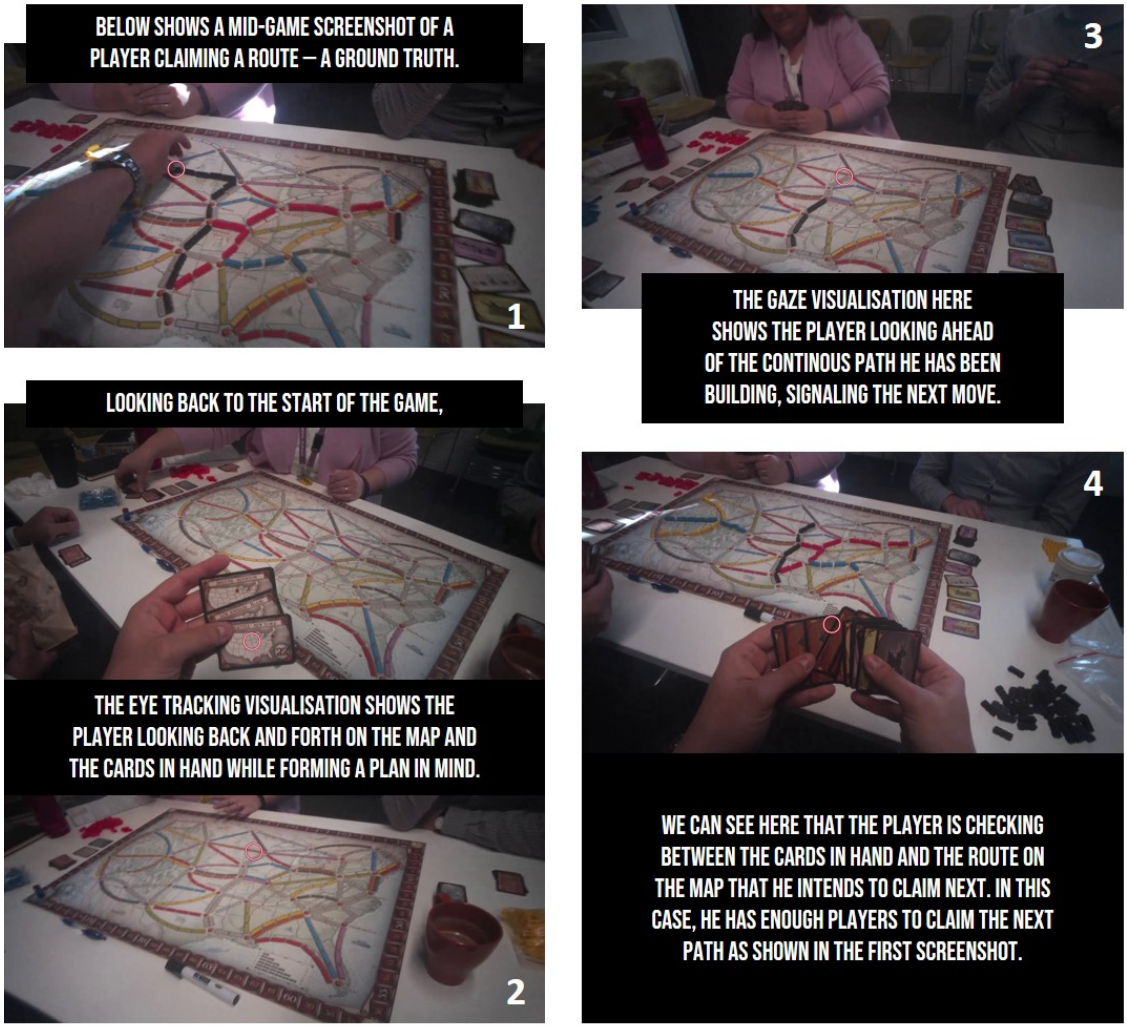


Figure 3.5: Example of early planning behaviour shown using gaze visualisation.

Behaviour 3: Inferring Intentions of Opponents

When the player waits for his turn, and has a plan in mind or has completed their current intended plans, the player will use the opportunity to observe the behaviour of other players, which includes their implicit cues such as where they are looking and other in-game actions such as what train card they have taken. As shown in Figure 3.6 below, the player looks at the gaze direction of an opponent across the table and infers where that opponent is looking on the board. The player then engages in *joint visual attention*, where he continues to follow where the opponent is looking on the board based on subtle changes in their gaze direction.

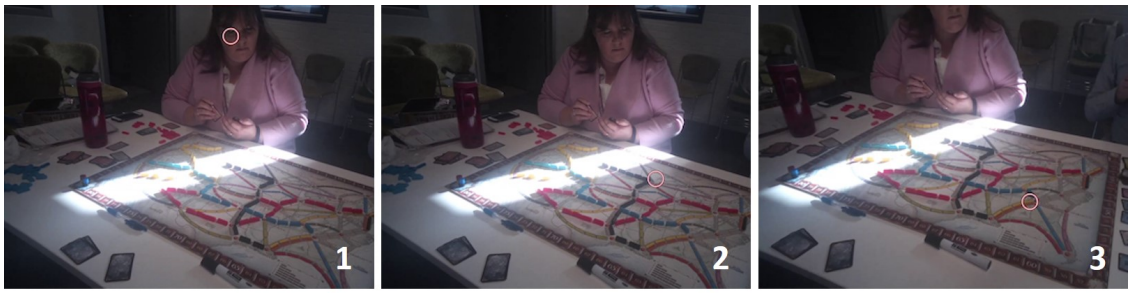


Figure 3.6: Example of how players track the gaze of one another during the game.

The outcomes from this feasibility study gave us the confidence to pursue the study of gaze-based intention recognition in our proposed setting and the use of the digital version of *Ticket to Ride* as a stimulus for eliciting gaze-based intentions. Beyond intentions, we note that there is a multitude of observations we can make from eye-tracking boardgame play [210]. In the next section, we outline the structure of the studies built around the behaviours shown in this study.

3.4 Structure of Studies

This section details the overall structure of the studies conducted in this thesis. In the introduction to the chapter, we presented an illustration of the overarching scenario (see Figure 3.1) which depicts a human and an artificial agent (represented as a humanoid robot) competing as a mixed-team involving a strategic turn-based board game—*Ticket to Ride*—against other teams. A closer look at the overarching scenario shows the different interactions and behaviours between actors that may naturally occur found in a real-world co-located competitive gameplay setting. As outlined by Pohl and Haumer, scenarios can be described at several levels for contextual

knowledge: *scenario level*—which provides the overall context including stakeholder goals and processes, and *interaction level*—which are more fine grained and describes interactions with the system or stakeholders [198]. Figure 3.7 below shows a breakdown of the whole scenario (scenario level) into manageable set of *interaction level* scenarios, in which we have leveraged to formulate the five studies conducted in the thesis to help answer the main RQ.

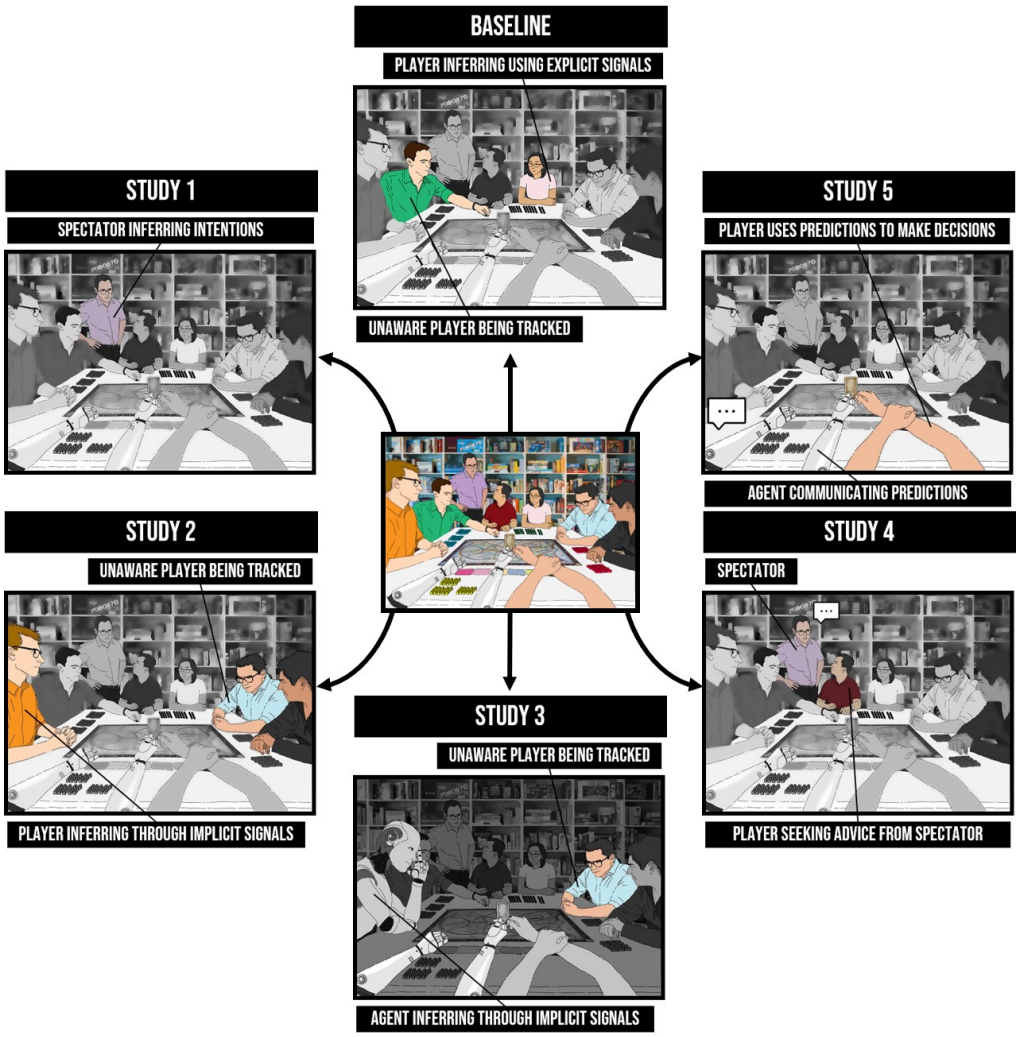


Figure 3.7: Breakdown of overarching envisioned scenario into *interaction-level scenarios*.

Below, we first provide a brief narrative for each of the interaction level scenarios shown in Figure 3.7 to avoid any misinterpretation. This is then followed by how each scenario relates to the research questions in the remainder of the chapter.

Baseline: The player in pink is predicting how the actions currently carried out by the

player in green will affect the state of the game and possibly her own strategy. The player in green is currently placing a game piece on the board during his turn, which 'explicitly' signals to other players his intentions to claim that part of the board, and along with past actions, a trajectory towards his goals. This scenario represents ***intention recognition using explicit actions***, and is typical of how players make predictions in such games.

Study 1: The spectator in this scenario has been silently watching both the game and the behaviour of the players from the start. The spectator knows the rules of the game well, and with no prescribed goals of his own, he is free to make his own inferences based on all the information he has gathered to predict how the game would unfold as it progresses. As such, the spectator can also observe where each player has looked and can use this complementary layer of information as part of his goal and intention recognition process.

Study 2: The player in orange is discreetly observing the nonverbal behaviours (e.g. gaze, actions, emotions) of an opposing player team across the table who are in the process of formulating and discussing their hidden strategy. Players (and their teammates) in games with hidden information can gain a significant advantage if they can correctly predict the plans of others. As the players have not become aware that they are being observed and that they are exhibiting nonverbal signals that give away information about their plans, this scenario represents ***intention recognition using implicit nonverbal signals***.

Study 3: In this scenario, an artificial agent (represented by a humanoid robot) replaces the human player in orange in Study 2. Similarly, the agent is discreetly observing the nonverbal behaviours of the opposing player team across the table as they form their plans in secret. The agent and by extension, its human teammate, can gain a significant advantage if they can uncover the plans of their opponents correctly, which will enable them to make better strategic decisions together as the game progresses. This scenario, therefore, raises the question of whether an agent can possess the capacity to make inferences using nonverbal signals as with the human player in orange in the previous study narrative.

Study 4: The player in red engages in a dialogue with the spectator who has observed how the game has unfolded as well as the behaviour of other players/teams since the start. Ideally, the spectator can provide recommendations on what the player should do next, but

this requires the player to communicate his plans. At the very least, the spectator should be able to communicate what the others plans are, or what they might do in their next turn. Either way, the spectator should be able to explain to the player how he came up with his inference. For instance, by describing the actions of an opposing team so far, or what he saw them looking at. In the previous scenario (Study 3), it raised the question of whether an artificial agent possesses the capacity to make inferences on human nonverbal signals. If the agent does possess this capacity, we can then start to question whether it has the capacity to not only effectively communicate its inferences but also whether it can explain them to the player when asked. Therefore, this scenario represents an instance of ***intention communication***, as any player or agent in the scenario should be able to explain its reasoning process once an inference has been made.

Study 5: This scenario describes an instance of a human-agent team collaborating, where the agent communicates its predictions of their opponents to the human player who then uses the information to make strategic decisions. In an ideal situation, the human and agent can negotiate on what would be the next move should be, especially when it differs between them. To keep scenario simple, the artificial agent in this scenario is tasked at gathering as much information as possible by observing the in-game actions and behavioural signals of the players in the opposing teams, forming a mental model of each opposing team (or player). The agent then communicates its inferences to the player and is able to explain how it came about them. The human player, on the other hand, will exercise judgement on what might be the best approach forward, and to do this effectively, the agent needs to communicate its awareness of the situation accurately and in an effective manner. The artificial agent, possessing the computational power and memory capacity to process all the information available, makes an ideal partner to assign the task of providing awareness by predicting and communicating the plans of its opponents.

By breaking down the background scenario into the five interaction level study scenarios along with a narrative of the actors involved and their interactions, it allows us to form a series of sub-research questions for each study, which can be found in the respective chapters—Chapters 4 to 8. For each study, we investigate a series of such interactions stemming from the scenario, where each study contributes to the design of an artificial agent, that when combined would

potentially allow the agent to possess the same capabilities as any other player in the scenario.

Figure 3.8 below provides an overview of the structure of the study chapters (Chapters 4-8). As shown, the studies are divided into three parts, each with a different focus. In the rest of the section, we provide the description and aims of each part and corresponding study, how their primary outcomes inform the subsequent studies, and how they align towards the final study, Study 5. Study 5 evaluates the experience of a human player that is assisted by an intention-aware artificial agent with the capability to perform gaze-based intention recognition.

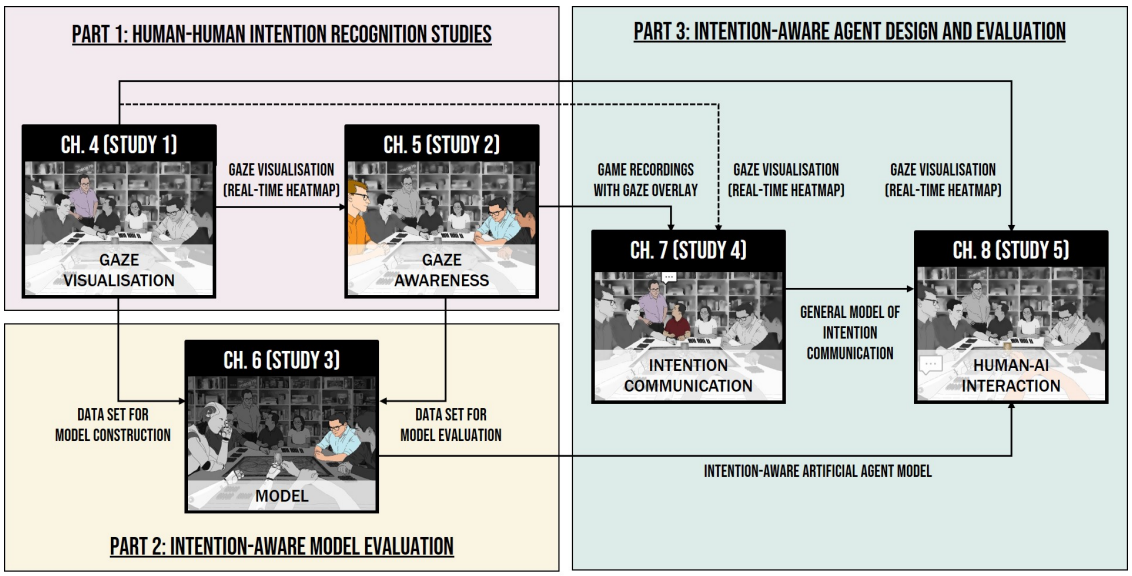


Figure 3.8: Overall Structure of Thesis Study Chapters.

PART 1: HUMAN-HUMAN INTENTION RECOGNITION STUDIES

To train an artificial agent to understand human gaze to perform intention recognition, we must first understand how humans portray and infer intentions from observing gaze behaviour. Hence, in this part, we evaluated how well humans can perform this skill when given the ability to see where others are looking (i.e. 'gaze awareness'). Simultaneously, we recorded a data set of gaze behaviour and ground truths as players form these intentions to evaluate whether an artificial agent can perform the same task in the next part. The first of the three sub-research questions, RQ1, builds on the *eye-mind hypothesis* [122], where we leverage the fact that the player will look at parts of the map they are interested in performing actions.

RQ1: Does gaze awareness improve the capacity of a human player to perform intention recognition in a competitive gameplay setting?

To answer this question, we designed two studies that explore the human capacity to infer intention, given the ability to 'see' the gaze of their opponent. Study 1 focuses on the visualisation aspects of gaze-based intention recognition, while Study 2 focuses on the effects of gaze visualisation during live gameplay in our chosen setting. To better inform the design of both studies, we first piloted them simultaneously before beginning Study 1. The description and findings of the pilot study are found in the introduction of Study 1 (Section 4.2).

Study 1: Real-Time Gaze Visualisation for Human-Human Intention Recognition

The first study in the thesis focuses on the ability of a human non-player (i.e. human spectator), to derive intentions from a range of visual representations of gaze. From our review of the literature, we found that gaze visualisation is the most common approach for utilising gaze input in technology-mediated human-human interaction. Gaze overlaid on a shared interface provides a complementary layer of nonverbal communication, especially beneficial in remote settings where the collaborators cannot see each other's eyes directly. However, users in previous studies have reported that its use can be *confusing* when there is a mismatch with other forms of communication [66] and *distracting*, due to the constantly moving live gaze visualisation on the interface [263]. Furthermore, our pilot study revealed similar findings that suggest that the *visualisation style* is highly dependent on the context, which also strongly influences the interaction. Gaze visualised in the form as a live 'red dot' augmented over the digital version of *Ticket to Ride*, may not be the most appropriate representation for an observer to perform gaze-based intention recognition effectively, as indicated in our pilot study. Hence, we formed sub-research question RQ1.1 to evaluate which visual representation of gaze would not only work best for performing gaze-based intention recognition but which could potentially suit a remote competitive gameplay setting—where only one player can see the gaze of the other (an unaware asymmetrical setup).

RQ1.1: How do we represent gaze to improve a human's capacity to perform intention recognition in a competitive gameplay setting?

This question is examined through a repeated-measures study that evaluates nine different visual representations of gaze in our chosen competitive digital game setting, each with different characteristics. The study empirically evaluates the ability of human spectators to infer a player's intentions for each visual representation. The results of the evaluation show that spectators have a remarkable ability to infer intent accurately using all nine visualisations, but that visualisations with certain characteristics were more comprehensible and more readily revealed the player's intent. The *real-time heatmap visualisation* was the most highly preferred by participants and the most effective in revealing intent, due to its ability to balance real-time gaze information with a persistent summary of recent gaze behaviour. This finding suggests that visualisations that show a small window of what a person is looking at work best, as the aggregation gives the observer time to process the information. Further, the effectiveness of the visualisation can be directly attributed to its style, as the observer can easily observe the areas that have gained more interest by noticing when the visualisation 'heats up'. In turn, the observer only requires a glance to know where the tracked player interests are on the interface, and therefore able to make inferences on what the players' intentions might be through the subsequent observations.

In comparison to visualisations that are directly mapped gaze coordinates (e.g. 'red dot' visualisation), the observer has to follow the gaze point and observe its behaviour constantly. An example of a behaviour is when the gaze point starts to 'hover' over an area, which is a potential signal for interest as it represents that the tracked player has looked longer at the area. As such, the need to constantly follow the gaze point not only consumes the observers time and attention but requires the observer to cognitively process what the gaze movements and patterns (fixation-saccade-fixation cycles) may signal, which together can be cognitively demanding.

Study 1 serves three purposes for the other subsequent studies in the thesis. First, the study serves as an initial investigation to validate our hypothesis that the sustained observation of gaze behaviour can lead to the recognition of human intentions. Therefore, the primary focus of the study is first to find out whether a human subject can perform the task of intention recognition, given the ability to observe gaze when displayed over the interface. For this reason, we employ the use of human spectators to simplify the task of intention recognition, focusing on their ability to interpreting intentions from visual representations of gaze without having to worry about managing their own strategy. In other words, a more concentrated test of interpretive

ability before introducing potential live distractions of live gameplay. Moreover, this allowed us to control for the underlying data, varying only the visualisation displayed to the participant.

Second, Study 1 aims to empirically investigate which visual representation of gaze would best enable players to perform intention recognition in the subsequent study. As mentioned, our pilot study revealed that using a 'red dot' visualisation suggests that might not be sufficient for aiding intention recognition. In the subsequent study, Study 2, players will have access to the gaze of their opponent throughout the game. Hence, it is important to equip the players with the best possible visualisation to better assist them in recognising the intentions of their opponent, as they would have to simultaneously manage their own strategy while keeping track of their opponent's as well. Further, it is also important for the players to infer intentions accurately as possible, as it will potentially give them a competitive advantage in the game. The knowledge of their opponent's strategic plans will potentially influence the decisions made by the player, especially if the inferred plans encroach on their own.

Third, the full game recordings used to generate the short clips for evaluation the gaze visualisations in this study resulted in an initial data set, consisting of both eye tracking data (gaze points) and ground truths (game actions). As shown in Figure 3.7, the data set informs the construction of our prototype artificial agent that incorporates gaze data into its planning. Hence, serving as an essential building block towards our goal of building a tangible agent artefact in the thesis.

Study 2: Effect of Gaze Awareness on Human-Human Intention Recognition

The evaluation in Study 1 showed that gaze, displayed as a *real-time heatmap visualisation* over the game interface, can improve a human's capacity to perform intention recognition in a competitive gameplay setting. In this lab-based study, we shift our focus from using *human spectators* to *human players* to study the effect of gaze awareness on intention recognition during live gameplay, forming sub-research question RQ1.2 that helps us to answer RQ1 fully.

> **RQ1.2:** What are the effects of gaze awareness for intention recognition in a competitive gameplay setting?

Gaze awareness, defined as 'knowing what someone is looking at' [175], has been shown to be an effective predictor of intention [10, 108]. In Study 2, we explore the relationship between gaze

and intention in a non-collaborative setting in contrast with existing works. In this study, we evaluate how well players can discern the overall strategic plans of their opponents, given the ability to see their gaze throughout the game using the real-time heatmap visualisation from the previous study. The study examines sub-research question RQ1.2 by comparing the results of three conditions using pairs of players: *no gaze visualisation*, *gaze visualisation of an unaware player*, and *gaze visualisation of an aware player*. We introduced the last condition to find out how players who are aware that their gaze is tracked will manipulate their gaze to keep their true intentions hidden. This study focuses on only using asymmetrical scenarios that form our the conditions tested to answer RQ1.2: (1) when a player is aware of where their opponent is looking, and (2) when a player gains awareness that their opponent is watching their gaze. The study extends the findings of the previous study in the following ways:

1. By using human players instead of human spectators, where players will make use of an informed gaze visualisation of their opponent during live gameplay, balancing the tasks of devising their own strategic plans with making inferences about their opponent.
2. By assessing the ability of human players to interpret and respond to opponents' gaze cues throughout a live game, with a particular interest in the ability to make predictions during the early stages game when little or no in-game actions have been taken.
3. By introducing a deception aspect where one player is aware that their gaze is being visualised, creating an opportunity for the player to deliberately manipulate their gaze behaviour to potentially mislead their opponents. Imperfect information add an element of mystery and uncertainty to a game. Imperfect information also invites treachery, trickery, and deception, and can be used as a design element in games meant to inspire mistrust and competition among players [213]. The experimental setup, where players are placed in separate observation rooms to simulate a remote game, allows us to study this aspect.

The overall results of the study showed that the awareness of gaze could further improve a player's capacity to recognise intentions of an opponent when displayed throughout the game, especially when compared to the baseline condition (i.e. no gaze visualisation) in which the player can only make inferences based on the opponent's in-game actions. As the opponent's gaze is displayed as soon as the game starts, the player can begin to make early predictions without having to wait for the opponent to explicit signal parts of their plans through in-game actions

such as by claiming a route or drawing certain cards. Players that were successful in making such inferences can potentially adjust their strategic plans accordingly to gain an advantage.

However, despite the increased capacity to perform intention recognition, players overall reported that they experienced a limited ability to keep track of their opponent's intentions and therefore unable to fully utilise the information that was theoretically available to them through the gaze visualisation. One possible explanation is that the opponent is continually evaluating alternate plans, and the player needs to determine the likelihood of the intentions as the game goes on. The player then has to continually make other strategic decisions in response, such as to differentiate the priorities of the opponent from their predicted intentions, determine how the intentions relate to one another, and how they might affect their own strategic plans. Hence, successfully managing the two plans that are continuously updated in a feedback loop and how they affect one another becomes a challenging task, resulted in no net gain in terms of win rates in the condition where the 'aware' player is given awareness of the 'unaware' player's gaze.

The study elicited similar reports on the *distracting* nature of using gaze visualisation when displayed over the interface as with previous studies (e.g. [65, 263]). The level of distraction experienced may be partly attributed to the visualisation style, either due to its representation or the fact that its a live visualisation. The choice of the real-time heatmap visualisation was informed by the previous study, in which it was the most accurately interpreted and second-most preferred visualisation style among nine formats compared by participants. Then again, participants in the previous study were spectators of a game rather than active players.

Further, the study also found that the capacity to perform intention recognition through gaze visualisation can be hampered when the opponent was aware that their gaze was being watched. Players were successful in devising deception strategies that reduced their opponent's prediction accuracy about their plans, but also found degrees of difficulty in maintaining them (in line with [205]). All of the deception strategies we observed were conceived independently by the players, without suggestions from the researchers, and the stated inspiration behind many of the strategies was the player's own experience of observing the gaze visualisation. Players were largely successful in converting their own experience into an accurate enough model of their opponent's thinking that they could implant a false belief about their strategy, or disrupt the prediction process in other ways. This finding demonstrates the application of *Theory of Mind*,

as the *ability to deceive* is considered one of its hallmarks [151]. As players describe how they attempted to deceive their opponents by manipulation their gaze (together with their actions), the study contributed a novel *taxonomy of gaze-based deception strategies* (Section 5.2).

PART 2: INTENTION-AWARE MODEL EVALUATION

In the previous part in which we focus on human-human intention recognition, the findings highlight the potential benefits of using gaze for improving intention recognition in our chosen setting. However, the findings from Study 2, in particular, indicated that the task of performing intention recognition using the combination of gaze and in-game actions while keeping track of the state of the game is cognitively demanding. This demand is further increased as the human players especially have to balance other tasks simultaneously, such as their own strategy. Therefore, the intention recognition task may be more suited to an artificial agent that theoretically possess both a higher memory capacity computationally ability to process the rich gaze data. A collaborative artificial agent as a teammate in the game can potentially alleviate a player's cognitive workload by performing the task of intention recognition on their behalf, which closely resonates with Licklider's *complementary and effective division of labour* prerequisite. The alleviated human player would then would only need to focus on forming the right strategy, but requires the agent to communicate the predictions effectively (to be investigated in Part 3).

At the same time, delegating the task to the agent enables the player to move away from the use of gaze visualisation, and therefore potentially reduce the confusing and distracting nature of its use while allowing the number of users to be scaled up from current scenarios (limited to two users). Due to these compelling reasons, the focus in this part is to evaluate whether an artificial agent can first improve its capacity to perform intention recognition provided by using gaze as *priors* in combination to observable in-game actions.

Study 3: Intention-Aware Gaze-Enabled Artificial Agent Evaluation

This study evaluates more broadly whether an artificial agent can improve its capacity to perform intention recognition with the introduction of gaze, regardless of the domain. Hence, the study compares the performance of base planning model is combined with and without gaze data, forming the second sub-research question RQ2.

RQ2: Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?

Study 3 employs the labelled data sets from both previous studies (as shown in Figure 3.8) to answer sub-research question RQ2. The data set from Study 1 served as a starting point for the construction of the agent model with promising results. The data set from Study 2 with multiple conditions allow us to evaluate the agent's capacity to perform intention recognition thoroughly, i.e., not only using natural gaze behaviours as *priors* but its robustness against deceptive gaze behaviours as well. The two comparative evaluations performed in this study are as follows:

1. First, we compared a model-based approach using only game-state data (*Model-Only*) to the same approach enhanced with gaze-based priors (*Gaze+Model*). The results in Study 2 showed that gaze awareness could improve a human player's capacity to recognise intentions when compared to a baseline, where no gaze visualisation was shown. Here, we make the same comparison to evaluate whether this is the same case for the agent.
2. Second, we compared the intention recognition performance results from the data set against the performance of the agent to determine if the agent outperforms the human player in recognising intentions. In Study 2, we found that players that are aware that their gaze is being watched would manipulate their gaze to hide their intentions, resulting in a lower intention recognition performance by the observing opponent. Hence, we compared the performance of the agent for both the natural and deceptive conditions.

As with human players, intention recognition is possible from just inferring from the observable in-game actions along with the state of the game, but can be improved with gaze information, especially when no in-game actions have taken place. The artificial agent works by using gaze data to build probability distributions over a set of possible intentions, which are then used as *priors* in the model-based intention recognition algorithm. The artificial agent optimises two features of eye movements, which are *fixation duration* (how long someone looks at an object) and *fixation count* (how many times someone looks at an object). Both fixations are reflective of the gaze behaviours displayed through the gaze visualisation in Study 2. However, as discussed, human players are subjected to confirmation bias when they strongly believe in a prediction.

The first comparison showed that although the Model-Only approach was reasonably successful

at predicting intentions, the model with the addition of gaze as *priors* resulted in more accurate predictions, which indicates an improved capacity to recognise intentions and therefore answers RQ2. Further, the combined approach resulted in more accurate predictions and earlier predictions. From our observations in Study 2, we found that players ignored large areas of the board (e.g. whole left side) when they could see that their opponent was not looking at it. Based on their comments, it is apparent that this ability to triage the board on a wide scale was a substantial part of the competitive benefit players derived from gaze visualisation, comparable at least to the more fine-grained prediction of specific routes. Similarly, the artificial agent with access to the gaze reduced its search space by 'ignoring' the irrelevant elements, which resulted in no additional computational execution cost despite the addition of gaze into its planning process.

The second comparison found that the agent outperforms the human players in performing intention recognition in both the natural and deceptive conditions from the previous study. We believe this is for two reasons. First, the agent has a sole task of providing the best plans possible using gaze as *priors*. Second, the agent can calculate features such a fixation counts and lengths, essentially filtering out the 'noise' from the natural constantly moving gaze point, which allows the accurate mapping on the areas of the game where the opponent is interested in. Hence, it is likely that players, despite displaying deceptive behaviours, their intentions to complete their goal remain and their respective true intentions are still 'leaked' across.

However, we note the current iteration of the agent model, being a prototype, does not perform well in all cases, especially when intentions are not sustained for long. We evaluate the performance of the agent based on the number of top plans it generates, in our case, the top 10 plans. This consideration is because their top plan might not be one that they are currently contemplating. The player may have potentially kept their primary goal in their mind while constantly seek out alternate plans. Nonetheless, the study demonstrates a working 'intention-aware' artificial agent with the capability to perform real-time gaze-based intention recognition for the next part.

PART 3: INTENTION-AWARE AGENT DESIGN AND EVALUATION

The previous two parts resulted in a working 'intention-aware' artificial agent prototype that can increase its intention recognition capacity when gaze is incorporated into its planning process. By employing the agent as a collaborator in our chosen scenario, it can potentially

alleviate the cognitively demanding task of intention recognition away from the human player. By assigning this task to the agent, it potentially allows the player to better focus on forming the best strategy possible given the information from the state of the game and the intentions from other players from the agent. The agent is therefore required to 'interface' with the player through communication protocols in an effective manner. Therefore, the third sub-research question, RQ3, raises a general question on the collaborative design of the agent in terms of providing support, given its proven and novel ability to perform gaze-based intention recognition:

> **RQ3:** How do we design gaze-based intention recognition systems for supporting human-agent collaboration?

This research question is examined through two studies, Study 4 and 5. Study 4 focuses on the communication ability of the agent, which will allow the agent to pass on its inferences to an assisted-player as a first step towards the agent's communication protocols. Study 5 serves two purposes for our broader aims of the thesis, to evaluate the experience of the assisted player interacting with a gaze-enabled intention-aware artificial agent during live gameplay and to further evaluate the agent and language models formed in the studies in the previous studies.

Study 4: Assistive Intention-Aware Agent Language Study

The literature dictates that communication is a crucial aspect of human-agent collaboration, leading to the development of well-established sub-fields in artificial intelligence, including natural language processing. This study investigates the communicative capabilities of the artificial agent, exploring how intentions could be communicated given the ability to perform gaze-based intention recognition to create an awareness of the situation, forming RQ3.1.

> **RQ3.1:** How should an artificial agent communicate intentions given the ability to perform gaze-based intention recognition?

To answer research question RQ3.1, we must first consider how an ideal agent would communicate intentions once recognised while taking into consideration the optimum level of support as different levels of artificial agent support can result in changes in cognitive workload, positively or negatively [43]. Hence, this study is focused on the *what*, *how* and *when* aspects of

intention communication. At this stage, we have chosen to a support a human collaborator by communicating the intentions of an opposing player as a first step, and not on providing recommendations as this requires the agent to know the plans of the human collaborator. This is because providing recommendations requires significant exploration on combining different mental models, and how best to communicate the intentions of the player to the agent. Hence, the ability to provide recommendations is currently beyond the scope of this thesis, and we will leave it for future work (Section 9.4).

Another important consideration is the artificial agent's explanation capability to support the human collaborator better, an important consideration for trust during collaboration [110, 216]. In real-world interaction, a human collaborator can produce an explanation on their inferences to a certain degree when asked, which makes it a crucial aspect of communication in collaborative settings. This capability is evident in the first two studies; where participants could verbally explain their reasoning process about their inferences when prompted. Therefore, we extend the intention communication model to include explanations to augment the agent-derived predictions for two primary reasons. First, users would often seek an explanation of its output when interacting with an agent to improve their own decision making [247]. Second, agents that provide explanations supports transparency and in turn, may improve trust in the system [100, 161, 174, 192, 249]. In our context, it is especially vital to include explanations as the agent can produce predictions that may not immediately be obvious to the player.

At the same time, the prototype agent evaluated in Study 3 supports transparency as it adopts a *white-box* approach, which allows us to interrogate the data or features to form an explanation for each prediction communicated. This transparency further allows us to adopt the Situation Awareness-Based Transparency Model (or SAT Model) [58] as a model 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. Using this model, we design the communication at two levels, at the middle level for enabling explanations and at the top level for communicating predictions.

For an explanation to be effective, it must be at the right level of detail [131]. 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, hence decreasing its effectiveness [192]. Aside from the *presentation format* (how to explain) and the *explanation content* (what to explain) [75, 131] of an explanation, it is also important to consider the application domain and the audience of the explanation [98], hence essential for us to situate our study in our chosen setting. Therefore, in this study, we use a variation of the Wizard of Oz technique to elicit how an assistant who can infer intentions using the combination of gaze and in-game actions would communicate the intentions to a teammate. In a typical Wizard of Oz study, a researcher secretly plays the role of the computer system while a participant interacts with it [132, 207]. In our study, participants were 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. The use of human elicitation was inspired by previous gaze-enabled human-agent interaction studies (e.g. RealTourist [202]). A similar approach has been used in the context of machine learning to 'bootstrap' a dialogue system on human-generated activity [208].

Using this approach, we collected how people communicate inferred intentions of tracked players and how they subsequently explain their reasoning process behind them. From this analysis, we formed a preliminary language model for communicating intentions that can be generalised to other settings at a behavioural level. Besides the knowledge of elements to include in the communication (i.e. spatial, temporal and quantitative [69]), the study finds that participants in their 'predictor-explainer' role also portray uncertainty when unsure about their predictions.

To sum up, Study 4 explored whether computers can communicate abstracted information derived from its ability to perform gaze-based intention recognition. The study determined the linguistic aspects of communicating inferences based on implicit gaze information and explicit in-game actions, and subsequently, how predictions can be explained in an assistive manner. This study further contributes to the rapidly growing field of Explainable AI (XAI), where we show that observable nonverbal behaviours can be adopted as a basis for explaining human intentions.

Study 5: Assistive Intention-Aware Agent Evaluation

The outcomes of previous studies converge at Study 5, which is the final study conducted in the thesis (illustrated in Figure 3.8). Figure 3.9 shows how the previous studies have addressed, to some degree, the different human-centred prerequisites presented in the previous chapter (Section 2.2). With reference to the figure, Study 1 and 2 helps enforces the importance of using nonverbal communication (i.e. gaze), with respect to intention recognition in a competitive gameplay setting. Study 2 and 3 demonstrates that gaze affords the ability to represent the mental states of a human player, in our case, the intentions of an opponent. In Study 2, players kept track of the intentions by annotating on a map of the game, while in Study 3, we produced a similar but improved graph where we can observe the likelihood of predictions through the agent from a planning perspective. In both cases, we can represent the mental state of the opponent. In Study 4, we address the language mismatch between the human and the agent, by investigating how humans communicate intentions using natural language, given the ability to predict them. Study 5 combines and simultaneous evaluates the two models arising from Study 3 and 4 to form components of an 'ideal intention-aware agent', which is the agent model and the language model, respectively. We designed both models were designed to work in tandem to improve the agent's capacity to interact with a human player in our chosen competitive gameplay setting.

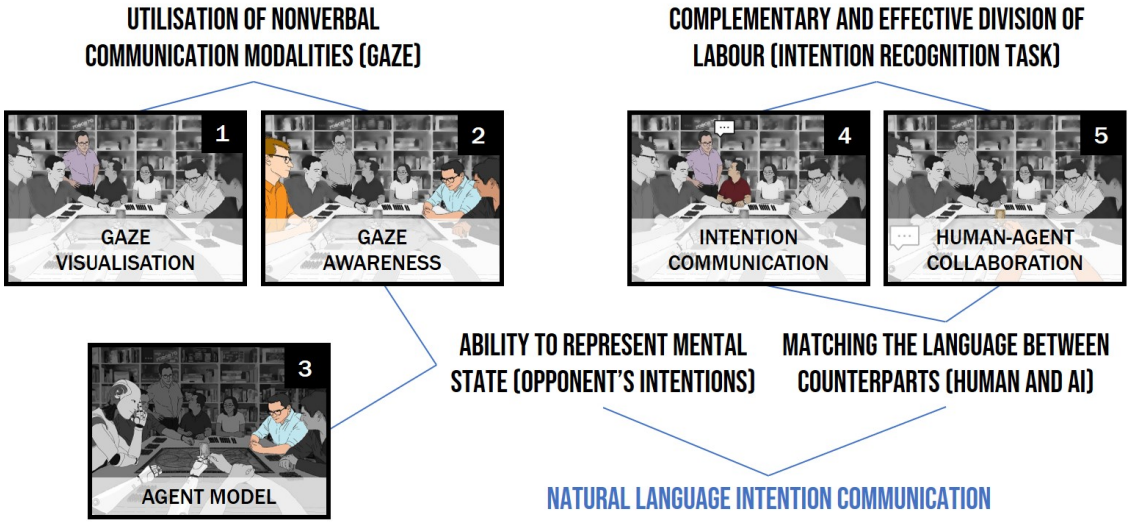


Figure 3.9: Relationship between human-centred prerequisites, studies and outcomes.

From Study 2, we acknowledged that humans are limited in their capability to infer and keep track of intentions, especially with the use of gaze due to its cognitively demanding nature. Gaze

provides a continuous stream of information about a player's interest in our scenario that makes it difficult to differentiate between its behaviours, which contributes to additional cognitive processing beyond keeping track of what has been looked at. For such reasons, we proposed that it might be suited to delegate the task of gaze-based intention recognition to the artificial agent, which is one way we can effectively divide the labour between the team. Hence, it has shaped the subsequent studies, particularly with Study 4, where we designed the study to focus on communicating intentions, in which the agent will eventually adopt. Therefore, with all the necessary components in place, we start to evaluate the live use of the artificial agent for collaboration in our setting, which brings us to sub-research question RQ3.2 below.

> **RQ3.2:** Does delegating the task of intention recognition to the agent improve the player's decision making in a competitive gameplay setting?

To answer this question, Study 5 focuses on the novel collaborative capacities of the intention-aware artificial agent, whereby the agent can provide situation awareness (with the capacity to explain them), and perform intention recognition on behalf of the human-agent team, which potentially reduces the cognitive load of the assisted human player. In theory, the increased situation awareness combined with reduced cognitive load would improve the player's decision making, as it would enable the player to focus on forming strategies with an increased awareness of the situation. This study focuses on the experience of players using an artificial agent that can project future states for effective strategy formulation afforded by the ability to create awareness through abstracted intention information, instead of the agent's performance to provide accurate and timely predictions on intentions. Hence, the study compared to the experience of users of using the gaze visualisation found in our first study as our baseline condition against the use of the agent, where we provide predictions at two levels, abstract (prediction only) and detailed (prediction with explanation). By comparing these two approaches, we can further determine the elements that are missing when we abstract intentions from the rich gaze information.

The findings 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 presented, suggesting that the manner in which the agent communicates its predictions requires further exploration. The predictions provided by the agent early in the

game allowed the participants to formulate better strategies. However, overall the agent neither impacted the players' performance nor decreased the cognitive workload as initially hypothesised. The game itself might have 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, most likely due to the reduced distraction (periodical abstracted information vs. continuous information), but also because there is no constantly moving gaze overlay. 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 in which the agent can determine what is the most relevant information to communicate, whereas 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. Hence, while textual representations of intention served as a good starting point for our work, it was insufficient to improve the player's performance (in terms of decision making). We look towards the use of visual explanation as our future direction, where agents can use multimodal (e.g. pointing, voice) or multimedia approaches (e.g. annotations, animations) to communicate intentions effectively.

Study 5 provides a first account of utilising a gaze-aware artificial agent in real-time for assistance in the context of strategic gameplay, which demonstrates the potential for building ubiquitous intention-aware systems. The findings of this study inform the development of agents with intention recognition capabilities from the perspective of human players. In Chapter 8, we discuss its design implications in terms of information presentation, level of detail, need for context awareness, usage of nonverbal signals, and making what the agent "sees" transparent.

In this chapter, we described our overall approach and provided the structure of our studies in detail. In the following chapters (Chapters 4-8), we present the details of each study, including the specific design and evaluation methods employed and a discussion of the respective findings.

The face is the mirror of the mind, and eyes without speaking confess the secrets of the heart.

– St. Jerome

4

Study 1: Real-Time Gaze Visualisations for Human-Human Intention Recognition

4.1 Introduction

This chapter presents Study 1, which was published as a peer-reviewed conference paper:

Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. 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* (CHI PLAY '17). ACM, New York, NY, USA, 541-552. DOI: https://doi.org/10.1145/3116595.3116624

The paper has been reformatted to the formatting style of the thesis as recommended by The University of Melbourne's *Thesis with Publication*[1] guidelines. The paper in its original published format is found in Appendix A.1.

The focus of this chapter is to evaluate the ability of humans subjects to predict intentions when shown a visual gaze representation of others, specifically, a tracked 'unaware' player. As humans are not accustomed to viewing live gaze visualisations, the chapter further evaluates which visualisation techniques would work best for improving their capacity to perform intention recognition, particularly in terms of visual characteristics and perceived usefulness. As described in Chapter 3, this study forms the first part of our human-human intention-recognition studies that helps us to answer RQ1. Therefore, this study is a necessary prerequisite to inform our subsequent study in the next chapter (Study 2) in terms of visualisation, where we evaluate the effects of gaze awareness of players in the same setting as shown in Figure 4.1 below.

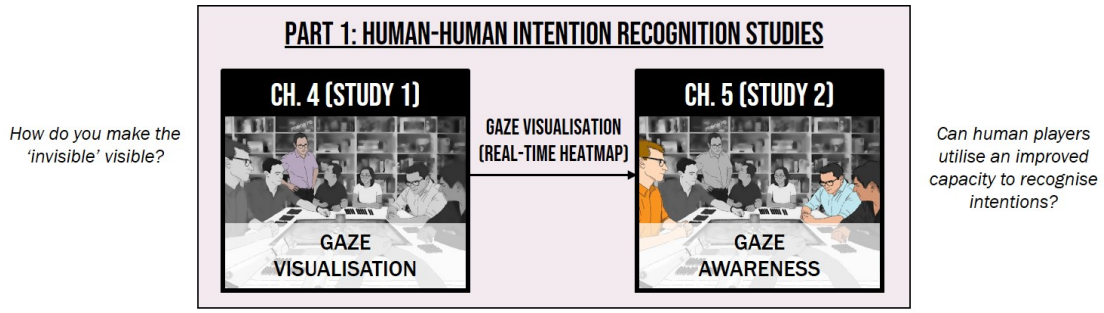


Figure 4.1: This study (Study 1) finds that gaze, represented as *real-time heatmap visualisation* works well for performing gaze-based intention recognition, which informs Study 2 in terms of visualisation.

[1] https://gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

From our review of the literature, we found that gaze visualisation is the most common approach for creating gaze awareness in technology-mediated human-human interaction. A visual overlay of gaze on a shared interface provides a complementary layer of nonverbal communication, especially beneficial in remote settings. Previous studies on gaze visualisation have that suggest that the *visualisation style* is highly dependent on the context. Hence, we performed a pilot study to examine the effects of using gaze visualisation in our chosen setting (see Section 4.2).

The pilot study indicates that gaze, represented as a live 'red dot' augmented over the digital version of *Ticket to Ride*, may not be the most appropriate or effective visual representation for increasing an observer's capacity to perform gaze-based intention recognition, warranting further exploration. Finding out which visualisation works best in our setting, especially in terms of revealing intentions is particularly important for our subsequent study (Study 2). Players will have access to the gaze of their opponent throughout the game and are required to simultaneously manage their own strategy while keeping track of their opponent's as well. Hence, a visualisation that would allow a player to infer the intentions of their opponent accurately in real-time with minimal effort would potentially give them an overall competitive advantage while playing the game. The knowledge of their opponent's strategic plans will potentially influence the decisions made by the player, especially if the inferred plans encroach on their own.

To help answer RQ1, we formed sub-research question RQ1.1 along with further sub-research questions to evaluate which visual representation of gaze would not only work best for performing gaze-based intention recognition but which could potentially suit a remote competitive gameplay setting—where only one player can see the gaze of the other (an unaware asymmetrical setup).

RQ1.1: How do we represent gaze to improve a human's capacity to perform intention recognition in a competitive gameplay setting?

- **RQ1.1.1** How well do human subjects infer intention from a range of visual representations of gaze in a competitive gameplay setting?
- **RQ1.1.2** What characteristics of gaze visualisation enable humans to perform better perform intention recognition in a competitive gameplay setting?
- **RQ1.1.3** Do gaze visualisation have an influence on predicting intent in a competitive setting?

This RQ1.1 is examined through this study, a repeated-measures study that evaluates nine different visual representations of gaze in our chosen competitive digital game setting, each with different characteristics (i.e. conspicuousness, visual information). The study empirically evaluates the ability of human spectators to infer a player's intentions for each visual representation.

The results of the evaluation using several measures show that human subjects (spectators) have a remarkable ability to accurately infer intent using all nine visualisations, but that visualisations with certain characteristics were more comprehensible and more readily revealed the player's intent. The *real-time heatmap visualisation* was the most preferred by participants and the most effective in revealing intentions, due to its ability to balance real-time gaze information with a persistent summary of recent gaze behaviour. This finding suggests that visualisations that show a small window of what a person is looking at work best, as the aggregation gives the observer time to process the information. Further, the effectiveness of the visualisation can be directly attributed to its style, as the observer can easily observe the areas that have gained more interest by noticing when the visualisation 'heats up'. In turn, the observer only requires a glance to know where the tracked player interests are on the interface, and therefore able to make inferences on what the players' intentions might be through the subsequent observations.

Besides informing which visual representation of gaze would work best in our subsequent study, this study serves as an initial investigation to validate our hypothesis that the sustained observation of gaze behaviour can lead to the recognition of human intentions. Therefore, a fundamental element of this study is first to find out whether a human subject can improve their capacity to perform intention recognition, given the ability to observe gaze when displayed over the interface. For this reason, we employ the use of human spectators to simplify the task of intention recognition, focusing on their ability to interpreting intentions from visual representations of gaze without having to worry about managing their own strategy. In other words, a more concentrated test of interpretive ability before introducing potential live distractions of live gameplay. Moreover, this approach using spectators and recorded games allowed us to control for the underlying data, varying only the visualisation displayed to the participant.

Before diving into the core section of this chapter, we included the details of the pilot study that we conducted to inform both Study 1 and 2 in the following section (Section 4.2). The section also includes the technical details of the *gaze visualisation system* employed in Study 1, 2 and 5.

4.2 Pilot Study

This pilot study aims to determine the suitability of *Ticket to Ride* further for our purposes. One of the factors for selecting this game is that it has a multiplayer digital version that can be played on and across a variety of devices (see Section 3.3), which partly contributes to its widespread play. Parts of this chapter has been published in a peer-reviewed work-in-progress paper: [189]. The paper included our initial formulation of the research idea, where we proposed the set of conditions for an initial study to understand the effect of gaze awareness on multiplayer.

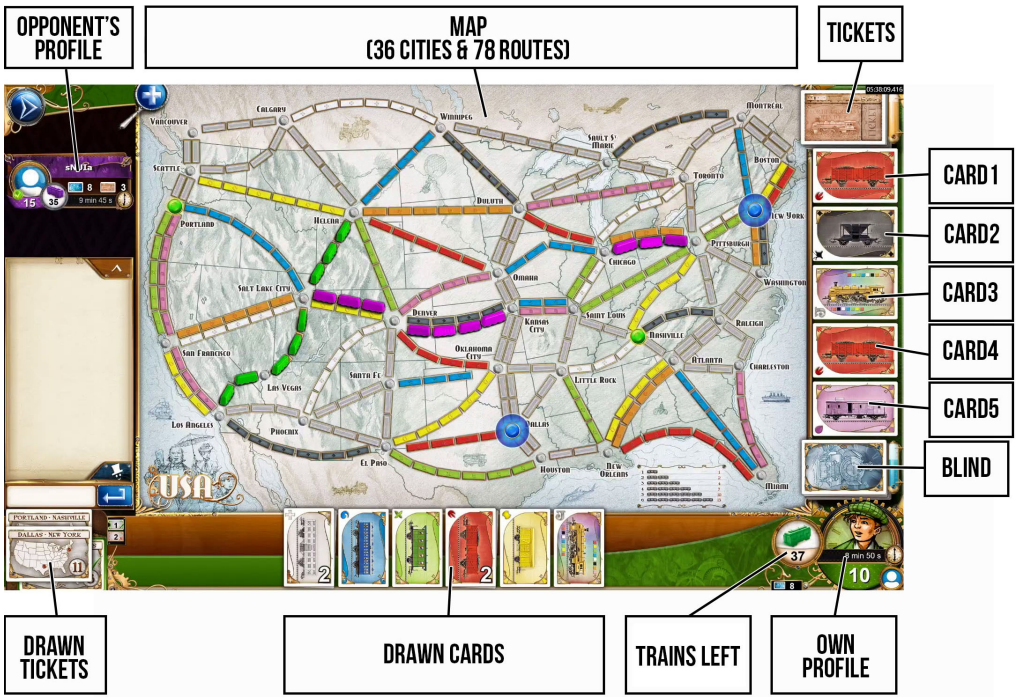


Figure 4.2: Screenshot of the digital desktop version *Ticket to Ride* board game.

Figure 4.2 above shows a labelled screenshot of the digital desktop version of *Ticket to Ride*. All players in the game see the same map and faced-up cards (Card 1-5) while the rest of the parts of the game will differ for each player. By adopting the desktop version instead of the physical board game, we can employ the use of screen-mounted (remote) eye trackers instead head-mounted (wearable) eye trackers. The fixed position of screen-mounted eye trackers affords better calibration and therefore allows us a better estimate the gaze of the players, both in terms of accuracy and precision. Moreover, this setup use makes it is much simpler in terms of display,

as we can use a 1:1 mapping to overlay gaze between players provided that the screen of both players have the same display resolution, and in terms of data logging for post-processing.

Figure 4.3 illustrates how the general setup looks like in a two-player game, from the perspective of each player. As shown in this example, only shared elements between the players are the map and face-up cards, and that the player on the left can 'see' the gaze of the player on the right, displayed here as a mapped 'red dot' over the game interface. Depending on the condition, the player on the right can either be aware or unaware that the player on the left can see a visualisation of their gaze. The players will see their own tickets (at the bottom left corner) and their own train cards in their hand (at the bottom). This means when the opponent's gaze hovers over these parts, the player is looking at their own hidden information and not of their opponent's. While it is possible to only display gaze in the shared areas of the interface, it would be interesting to see how players make sense of gaze behaviours through visualisation, such as comparing between the cards in their hand (or their drawn ticket cards) with the routes and areas on the map, reflective of the behaviours exhibited in our feasibility study (see Section 3.3.3).

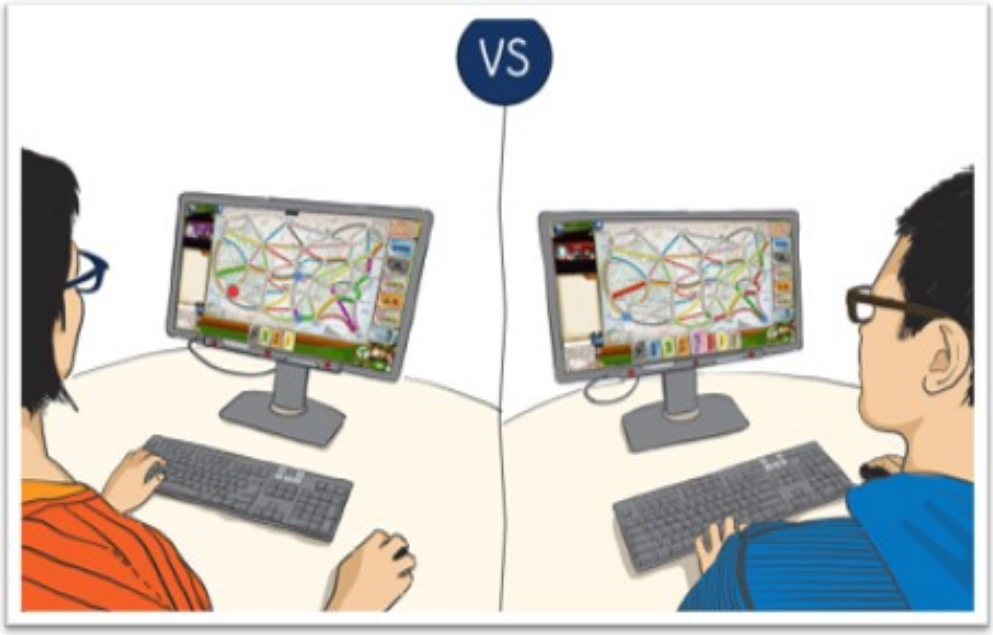


Figure 4.3: Gaze visualisation example (using a red dot for visualisation).

In a typical co-located gameplay setting, players can make inferences by observing each other

nonverbal cues, for instance, the direction of another player's gaze, as shown in our feasibility study. However, gaze is typically absent in remote multiplayer games. However, by introducing gaze using eye trackers in such settings, we can not only bring back some of the social aspects of gaze but can add a playful dimension. For instance, a remote multiplayer setting allows us to manipulate the awareness of the player in two ways: whether a player is aware of their opponent's gaze (gaze awareness) and (2) whether a player knows that their gaze is being tracked by their opponent (awareness of being tracked or observed). Hence, we formed a set of four initial conditions to test in this pilot study—by manipulating what the player *sees* and what the player *knows*. These conditions are illustrated in Figure 4.4 below, and explained accordingly.

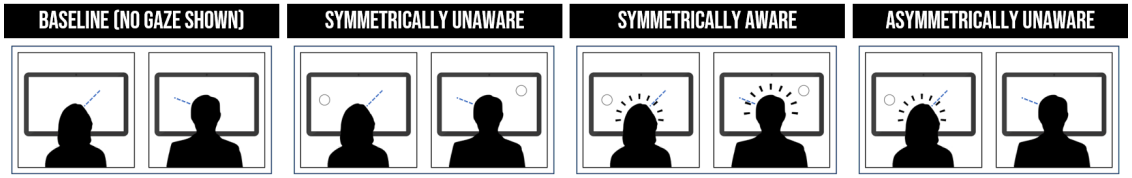


Figure 4.4: Pilot study conditions. The circles represent a gaze visualisation overlay on the screen.

1. **Baseline** (No Gaze Shown): This condition serves as a baseline condition, where the gaze of their opponent is not visible to either player and that both players know this fact.
2. **Symmetrically Unaware**: Both players can see the gaze of the other player but believe that they are the only player with this ability/power. In this condition, we are interested in whether players will be able to discover that the other player can see their gaze, for example, by changes in their gaze patterns or changes game strategy.
3. **Symmetrically Aware**: Players can see the gaze of the other player and are aware of this fact. This condition is analogous to the co-located version, where players can access the nonverbal cues of others at all times throughout the game and is aware that the other players can do so.
4. **Asymmetrically Unaware**: Only one player is aware of their opponent's gaze and at the same time, and aware that the other player does not have this ability. In this condition, we are interested in what happens when the game is tipped in favour of a single player who can see the gaze point (or visualisation) of an 'unaware' opponent.

With reference to Figure 3.7 in the previous chapter, we can imagine how these conditions are possible interactions between real-world players in a co-located competitive gameplay setting.

To visualise the gaze of players, we built a custom networked system that can read and overlay gaze over any application, in our case, the digital desktop version of the *Ticket to Ride* game. Figure 4.5 shows the system architecture, which consists of three main components:

(1) the **main application** (with networking capabilities) that reads and broadcasts the gaze from the eye tracker,
(2) a **transparent window** (see-through, click-through) that displays the gaze visualisations,
(3) a **settings window** to connect the application to another instance over the network, and to change gaze visualisation and its respective the parameters if needed.

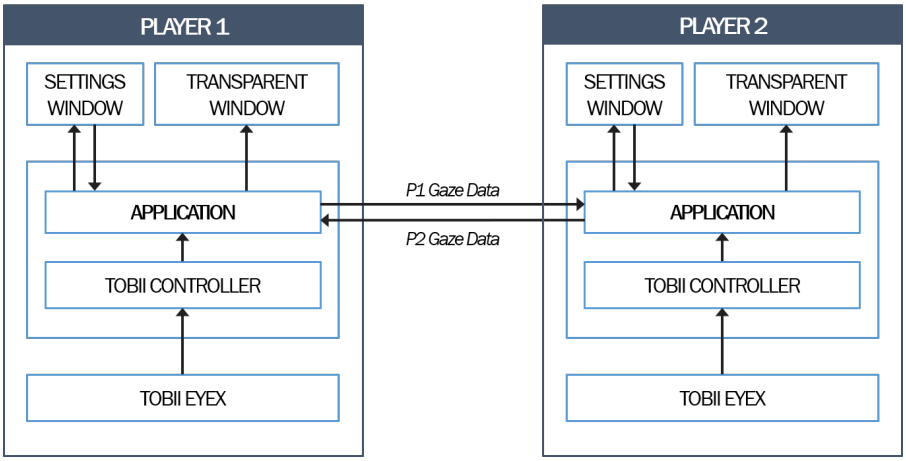


Figure 4.5: Gaze Visualisation System Architecture.

The gaze visualisation system is built using Windows Presentation Foundation (WPF) in C#, and uses the Tobii Core SDK[2] to obtain the data from the eye tracker. The Tobii Controller from the Tobii Eye Tracking Core Software[3] installed on the system runs as a background service, which allows multiple applications to access the data from the eye tracker concurrently. We added network capabilities to the system using the OSC communication protocol[4] to enable the application to concurrently receive and stream the gaze data across the network to and from a connected instance of the application. As shown in Figure 4.5, Player 1's gaze is streamed to Player 2's instance of the system and vice versa. The system allows us to toggle between the different data streams and modes from the Tobii SDK: gaze data stream (lightly filtered or unfiltered) and fixation data stream (slow or sensitive).

[2]https://developer.tobii.com/consumer-eye-trackers/core-sdk/
[3]https://gaming.tobii.com/getstarted/
[4]http://opensoundcontrol.org/introduction-osc

The transparent window for displaying the gaze visualisations is both *see-through* and *click-through*, which allows the players to play the game without any changes. Further, the transparent window is capable of displaying the gaze of one or both players simultaneously, meaning that we can toggle the gaze visualisation of one player to be visualised by themselves or by another user, or both at the same time. Additionally, in the settings window, we can change the gaze visualisation on the transparent window (including the parameters), including enabling different gaze visualisations on the same gaze data stream. For this pilot study, we simply visualise the lightly filtered gaze data stream using a 'red dot' (30-pixels radius, 1:1 mapping). From Study 2 onwards, we upgraded the model of the eye tracker from Tobii EyeX (30Hz) to Tobii 4C[5] (90Hz).

Study Procedure & Aims

We installed the gaze visualisation system, eye trackers and *Ticket to Ride* on two computers located in two separate observation room at the Usability Lab at The University of Melbourne and tested them accordingly (identical setups). We then recruited 8 players (4 pairs) to take part in our pilot study. Depending on the condition, a player can be the *naive player* (not aware that their gaze is being tracked) or the *aware player* (aware that their gaze is being tracked), as illustrated in Figure 4.4. Players were also sent the rules of the game upon recruitment.

We placed players in the separate observation rooms along with a co-experimenter when they arrived. After a short briefing, players first performed the default calibration procedure of the eye tracking device and were instructed to play the game's interactive tutorial until both players were satisfied that they understood the rules and mechanics of the game (for up to 10 minutes). For all pairs, players would then play two subsequent rounds, with the first round being a standard game (baseline condition) to get familiar with playing the game and its rules.

For the second round, the first three pairs played tested one of the three non-baseline conditions, i.e. *symmetrically unaware*, *symmetrically aware*, and *asymmetrical unaware* respectively. The fourth pair played the *asymmetrical unaware* condition as we wanted to further test whether the condition favours the *aware player*. For both players, we recorded the display using *Open Broadcaster Software (OBS) Studio*[6] and logged the gaze data for post-processing using our system.

[5]https://gaming.tobii.com/tobii-eye-tracker-4c/
[6]https://obsproject.com/download

As mentioned, this pilot study is aimed at determining the suitability of *Ticket to Ride*. To elaborate, we wanted to find out the digital version of the game would suit the needs for the overall work, particularly with Study 1 and 2. To some degree, the study is designed to further motivate Study 1 in terms of exploring gaze visualisation, as the design of visual representations in graphical user interfaces is highly dependent on their context of use. Hence, we evaluated a rudimentary visual representation of gaze itself in which we employed a small red circle or 'dot' (30-pixels radius) similar to those used in prior work as a starting point in our research.

In terms of Study 2, the pilot study tests whether players could portray/leak intentions to an observer through visualisations and at the same time, infer intentions using gaze visualisation. Further, we wanted to test the overall effects of introducing gaze visualisation over a live competitive game, which will help us to refine the study conditions and procedure for Study 2. Hence, upon completion of the study players, we sat both players together and asked them individually about their gameplay experience. In particular, we wanted asked what they liked/disliked, their feelings when their gaze was exposed, whether the knowledge of their opponent's gaze gave them an advantage and whether they could discern the intentions of their opponent. As we adopted an open-ended interview approach, participants were also free to express their thoughts and provide any suggestions and improvements. There was no time limit imposed on the games, which helped us to initially gauge how much time is needed for each session for Study 2.

Study Findings

From our interviews, we found several reoccurring themes among the players in which they primarily spoke about their experience with the added gaze visualisation. Gaze visualisation was the novel factor for 6 out of 8 players[7], hence formed the bulk of our preliminary findings.

Players noted that the ability to see the gaze of their opponent allowed them to gain insights into their opponent's intentions, which gave them the feeling of having 'an edge' over their opponent. These players also noticed that gaze visualisation was much more useful at the early to mid-stages of the game. At these stages, there was little information present on the board (claimed routes), and the gaze of their opponent provided some indication of where their opponent was going to claim. Players also reported a higher level of social presence experienced in line with previous

[7]Two players were 'naive' and were not subjected to gaze visualisation in the *asymmetrical unaware* condition.

studies that investigated the effect shared gaze visualisation in multiplayer games [147, 168].

However, all players who could see the gaze of their opponent mentioned it was difficult to follow and was 'distracting' as the gaze point was constantly moving (similar feedback in previous studies that employed live shared gaze visualisation). The downside of using a visualisation that is directly mapped to the gaze coordinates (i.e. without any processing), is that the player needs to constantly follow it to know where the opponent has looked and is currently looking.

Players who could see the gaze of their opponent also reported an increase of perceived mental workload, as players wanted to track of their opponent's intentions but same time has to form and play out their own strategies. This led them to have some conflicts on strategy formulation, as they had to choose between playing their own strategies and disrupting their opponent's. In one extreme case, the player was confused between their own strategy and the opponent's strategy, and ended up claiming the wrong routes. This increase in perceived mental workload is further exacerbated when they are also aware that their opponent was watching their gaze (*symmetrically aware*), as simultaneously attempted to manipulate their gaze to keep their intentions hidden.

Moreover, in the symmetrical condition where both players can see the gaze of their opponent, we observe a level of *gaze following*. This means that players constantly followed the gaze of each other, resulting in an undesired feedback loop. This behaviour led to confusion on whether they were actually tracking their opponent's natural gaze behaviour and by extension, their intentions or a reflection or their own. Therefore, from this pilot study, we concluded that symmetrical conditions are not suitable for the purposes of our work, particularly for Study 2. With regards to the *asymmetrically unaware* condition tested, we believe this setting is a better reflection in terms of human-agent collaboration and better aligns to the aims of the thesis. In real-world settings, an artificial agent can potentially 'see' the gaze of a human but not vice versa, and that the human may or may not be aware that their gaze is being tracked depending on the setting. In our case, players did not notice that their gaze was being tracked and this led to natural gaze behaviour displayed to the aware player. Lastly, to train an artificial agent to make the same inferences, a good starting point is to obtain a data set with natural gaze behaviours.

In the following section, we present Study 1, in which we evaluated the real-time visual representations of gaze for inferring intentions in our setting, motivated by the inadequacies found.

4.3 Published Paper: *Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games*

Start of Published Paper

Abstract

In this paper, we investigate nine different visual representations of gaze in a competitive digital game setting. We evaluate the ability of spectators to infer a player's intentions in the game for each visual representation. Our results show that spectators have a remarkable ability to infer intent accurately using all nine visualisations, but that visualisations with certain characteristics were more comprehensible and more readily revealed the player's intent. The real-time *Heatmap* visualisation was the most highly preferred by participants and the most effective in revealing intent, due to its ability to balance real-time gaze information with a persistent summary of recent gaze behaviour. Our findings show that eye-tracking visualisation can enable playful interactions in competitive games based on players' ability to interpret opponents' attention and intention through gaze information.

Introduction

Players' nonverbal signals are an important element in many competitive board and card games. By monitoring an opponent's body language, facial expressions and gaze direction, a player can infer their next moves or call their bluffs. Gaze is a strong signal of a person's area of attention and interest [85], and can give away unintended information—what psychology researchers call 'nonverbal leakage' [77]. Gaze is equally important in collaborative settings, as a cue for collaborators to monitor each other's attention and coordinate their actions [133]. Despite its importance in co-located settings, shared gaze awareness is typically absent in remote multiplayer games. In this paper, we demonstrate an opportunity to bring shared gaze awareness into this setting by comparing nine different ways of visualising gaze within a digital multiplayer version of a popular board game—*Ticket to Ride*[8].

[8] https://www.daysofwonder.com/online/en/t2r/

Figure 4.6: An opponent's gaze represented as a real-time heatmap visualisation over the *Ticket to Ride* game, showing interest in the east coast of the North America.

The advantages in observing a partner's gaze in a *collaborative* task are clear, leading researchers to build multiple systems where users' gaze points are streamed over a network to improve coordination [37]. However, the opportunities for gaze in *competitive* settings are more nuanced. Recent studies have explored these possibilities by using visualisations of players' gaze to open up novel gameplay experiences and increase social presence [147, 189]. Normally in remote multiplayer games, a player's only clues for predicting their opponents' current and future plans are the in-game actions they have previously taken; players do not have access to the bodily cues (such as gaze) that enrich the experience of games played around a table or in a shared space. When given information about an opponent's gaze in a networked game, a player can build hypotheses about what their opponent is thinking through observation of the areas or game elements that receive the most visual attention. This creates further opportunities for a player to deliberately manipulate their own gaze behaviour to deceive an opponent about their plans and goals. This contrast between the tendency for observed gaze to reveal intent and the ability of a player to use their eye movements deceptively makes shared gaze awareness a rich resource for game design.

However, studies on shared gaze visualisations for various applications have found that the introduction of shared gaze awareness can be *disruptive*, often being referred to as 'distracting'

and 'confusing' [66, 147]. We hypothesise that this is caused by two primary problems with gaze visualisation for digital interfaces. First, the way in which gaze has been visually represented in past works may be inadequate or ill-suited for the task and setting. This is expected, as the design of visual representations in graphical user interfaces is highly dependent on their context of use. This problem is exacerbated by an inherent problem with gaze: gaze is 'always on', which makes designing gaze representations difficult as the technology cannot distinguish between focused scrutiny and more absent-minded looking.

Second, humans are not accustomed to interpreting visual representations of gaze, as the focal point of gaze is 'invisible' in normal interpersonal interaction. Gaze information can be 'noisy' due to the jittery movements and the way the eyes work (fixation-saccade-fixation cycle), and this is reflected in the representations. Gaze data is difficult to interpret and often misunderstood [66], as even trained professionals in eye-tracking evaluations make use of a range of gaze visualisations to make sense of the same collected data. Further, as gaze is fast-moving and never entirely still, any real-time visualisation overlaid on a user interface adds a highly distracting element that may interfere with the viewer's own thought process.

To effectively use gaze awareness in a game, the designer must understand how the choice of gaze visualisation will affect players' understanding and enjoyment. Inspired by how players observe the gaze of one another in tabletop boardgames, we selected a popular board game—*Ticket to Ride*—for the purposes of our research. Using recorded clips from eye-tracked gameplay, we assessed nine gaze representations in three different stages of the game (early, mid, late). This paper serves as an initial novel groundwork for determining the characteristics of gaze representations that allow human players to make inferences about other players' intentions and into derivation of potential insights into their strategies (goals and plans).

Related Work

In this section, we summarise related works that require a user to interpret gaze representations of another, which falls into two key areas: research on *shared gaze*, and works that compare and interpret the use of gaze representations. We have chosen to exclude works related to eye-tracking evaluations in which gaze data is interpreted by trained professionals, as our interest is specifically on the ability of non-specialist users to interpret visual gaze representations.

Shared Gaze

Shared gaze systems are those that give users the ability to see the gaze of a partner, typically to improve communication and coordination between pairs in collaborative problem-solving tasks. Here, the eyes are typically not directly visible, and the addition of a gaze visualisation provides a complementary layer of nonverbal communication. Knowing where someone is looking can provide rich information, from inference about their intentions to clues about their current cognitive activity (based on behaviours such as scanning, focused interest in an object, and repeated comparison of different objects) [231]. With this knowledge, much research in HCI (Human-Computer Interaction) has used gaze in collaborative settings to foster coordination. For instance, Stein and Brennan [231] found an improvement in problem-solving task performance when another person's gaze was used as a cue. Gaze has been shown to be effective in co-located settings, e.g. when multiple users perform a visual search task on a large display [263]. A common approach is to present a user's gaze as a cursor overlaid on the shared visual display space, to create an awareness of that user's focus area [59, 231].

Gaze has shown promise when used implicitly with other modalities in remote communication. Qvarfordt and Zhai applied gaze in a dialogue system, where a remote assistant is able to detect the interest of a remote user through their gaze patterns while conversing with them (remote user is unaware that their gaze is being tracked) [202]. Tracking gaze behaviour can serve as a natural deictic pointer and thus reveal interest in a natural way. Brennan et al. have demonstrated that shared gaze can be more efficient than speech for the rapid communication of spatial information (spatial referencing) [37].

Despite the numerous benefits mentioned, gaze as an input suffers from inherent problems is that gaze is 'always on', related to the 'Midas Touch' problem that leads to accidental triggering of gaze-enabled interface widgets [116]. To combat this, gaze is often paired with another modality, such as voice interaction; however, this can also lead to confusion when there is a mismatch between gaze and voice inputs [66]. However, in competitive settings, this ambiguity can work as a resource for game design, revealing a glimpse into players' intentions without giving it away entirely.

Gaze Representation

How gaze information is represented strongly influences gaze interaction [54]. We reviewed two recent works in HCI that compare and contrast gaze representations to gather design implications for our study. Zhang et al. evaluate gaze representations for use in co-located collaboration, where two users share a screen and two gaze points are represented [263]. From a review of prior works, the authors selected four different gaze visualisation formats for testing—*Cursor*, *Trajectory*, *Highlight* and *Spotlight*—of which participants preferred highlight and spotlight. The authors determined that attaining a balance between *visibility* and *distraction* was one of the biggest challenges when designing and selecting gaze visualisations.

In another work, Li et al. designed and evaluated two gaze representations aimed at improving the coordination between pairs of users [153]: *zoom focus* and *gaze trail*. Both representations used fixations processed from the gaze stream; the former aimed to show the precise location of a fixation, while the latter aimed to show the most recent gaze locations. *Zoom focus* was found to be cognitively demanding and *gaze trail* was found to have excessive information. The authors conclude that when designing gaze representation, one should consider *distraction* and *visual information*.

Games Context and Summary

Gaze is an emerging input in computer games, with a focus on using a player's own gaze as a direct control mechanism to enable novel playful experiences [240, 243]. There is growing interest in implementing gaze as a mechanic for multiplayer games, in both cooperative [168, 197] and competitive [147, 189] settings. While the use of gaze in cooperative gameplay can directly leverage the findings of 'shared gaze' literature (e.g. [197]), competitive gameplay presents a different design challenge, as players have opposing objectives and are more likely to see different information on screen. We seek to find a gaze representation for use in competitive games that maintains the game balance, giving neither player a disproportionate edge over their opponent while not being disruptive to gameplay.

The existing literature makes clear that there is a need to balance *visibility*, *distraction* and *visual information*. Furthermore, we note that the problem with a majority of gaze representations is that they are designed to be used in post-processing of gaze data. Therefore, there is a need to design

and evaluate novel gaze representations, such as that presented by Li et al. [153], to cater for settings where gaze input is not commonly used. Before we can design novel gaze representations, we need to understand the intrinsic characteristics of existing representations in relation to the context of use. In the next section, we present a variety of visual gaze representations that we have selected for use in our study.

Real-Time Gaze Representations

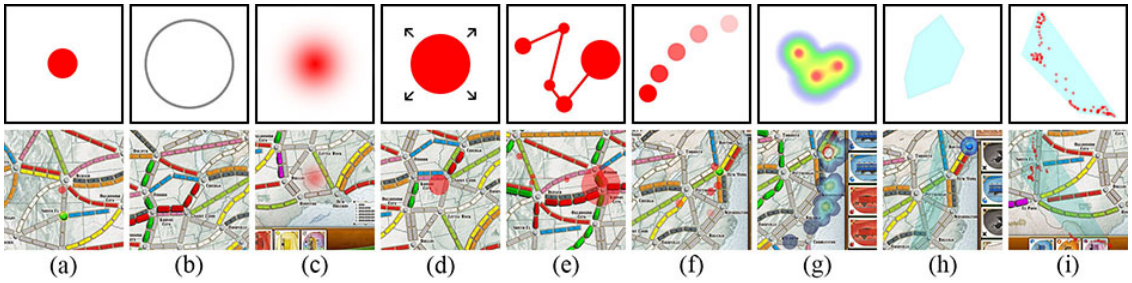


Figure 4.7: Top: The nine real-time gaze representations selected. L-R: (a) Dot, (b) Cursor, (c) Spotlight, (d) Fixation, (e) Scanpath, (f) Fixation Trail, (g) Heatmap, (h) Convex Hull and (i) Bee Swarm. **Bottom**: How each representation appears over the *Ticket to Ride* game.

Figure 4.7 shows the nine visual gaze representations which were evaluated in this study. For each representation, we provide an account of its current applications, its characteristics and the parameters for our use in the study. Most of the representations are common in existing in eye-tracking software, such as Tobii Pro Studio[9]. These representations are primarily used for gaze data analysis and to replay collected gaze data. We also considered a wide selection of representations from works that: (1) survey gaze representations (e.g. [31, 163]), (2) evaluate gaze representations (e.g. [263]) and (3) use gaze for interaction (e.g. [197]). In making our selection, we required the representations to be clearly distinguishable from one another. The selection was based on the characteristics *conspicuousness* and *visual information* (see Table 4.1). We chose the term *conspicuousness* (rather than 'visibility') to highlight the degree to which the representations can be used to display and draw attention, beyond just how well the viewer is able to see the representation. *Visual information* refers to the amount of information the representation provides. This is measured by the amount of aggregated gaze data used to generate the representation in addition to its overall size. We further sub-categorised the selected

[9]http://www.tobiipro.com

representations by their characteristics into three equally-sized subgroups: (1) Point-based, (2) Trajectory-based and (3) Area of Interest (AOI)-based.

Representation	*Conspicuousness*	*Visual Information*	*Category*
Dot	Low	High	Point
Cursor	Low	Low	Point
Spotlight	Low	Moderate	Point
Fixation	High	Low	Trajectory
Scanpath	High	Moderate	Trajectory
Fixation Trail	Moderate	Moderate	Trajectory
Heatmap	Moderate	High	AOI
Convex Hull	Moderate	Low	AOI
Bee Swarm	High	High	AOI

Table 4.1: Real-time gaze representations in relation to conspicuousness and visual information.

Point-Based Representations

For this category, we first selected two types of gaze representations previously examined by Zhang et al. [263]: *cursor* and *spotlight*. We then added the commonly-used *dot* representation as a contrast to the first two in terms of precision and visual information. The point-based representations are mapped directly to the real-time gaze coordinates from the eye tracker. To reduce the jittery motions that naturally arise from eye movements, we applied a weighted moving average filter [226] on top of the eye tracker's already lightly filtered data to smooth out the motion. This gave all three representations a hovering effect when a user's eyes moved across the screen and reduced distracting jitter during gaze fixations.

Dot: This representation appears as a translucent red dot, akin to a laser pointer beam or, at larger sizes, a torch beam (Figure 4.7a). It has commonly been adopted for use in gaze-based interaction as a deictic pointer and has been demonstrated to be an effective referential pointer to facilitate coordination in remote collaboration [11]. This representation allows a viewer to see gaze with high precision but can be potentially distracting. Our implementation used a 0.5 opacity with a radius of 30-px.

Cursor: This representation is a large ring with a transparent centre (Figure 4.7b). Its use has been demonstrated by Zhang et al. [263] and in GazeArchers [197], a game in which two players use gaze and touch simultaneously on a shared surface. Further, Tobii divisions has implemented

a similar representation in many of their recent products (Tobii Gaming[10] *Gaze Trace*, Tobii Pro *Gaze Share*[11]). Tobii Pro also uses this representation for their wearable eye tracker controller to view the gaze of a user in the real world in real-time. This representation is equivalent to the size of the foveal vision which is about the size of a thumb at arm's length, making explicit the confidence range of the gaze estimation. We matched the size of this representation to default size of Tobii's *Gaze Trace* (150-pixel radius) and coloured the outline in grey to give this a low visibility. This representation provides more visual information than the previous representation (*Dot*) as it allows users to see what is enclosed in the ring as opposed to a specific point.

Spotlight: Figure 4.7c depicts this representation where the gaze point is shown as a red circle with high opacity in the centre and decreasing opacity towards the edges. This was the most widely preferred representation in Zhang et al.'s evaluation [263]. It was noted to be more subtle and less distracting due to the reduced overall visibility and conspicuousness. We modified the parameters to suit our study, reducing its overall size to a radius of 90 pixels (30 pixels at the highest opacity) and changing the colour to red, as white blended into the light background of the map area of the *Ticket to Ride* game board.

Trajectory-Based Representations

The representations in this category use the sensitive fixation stream provided by the eye tracker's SDK to reveal the sequence of points at which the player has recently looked, leaving behind a short history of the player's gaze. We initially drew from professional eye tracking usability studies that use animated representations to replay gaze data. In our review, we found similar representations being used in eye tracking research. For example, Qvarfordt and Zhai use a scanpath representation to infer dynamic eye movements in their study [202]. We also found proposed representations aimed to enhance remote collaboration, demonstrating the importance of assessing representations that fall into this category [153]. A red fill is used for representations in this category, with each fixation presented as a translucent ellipse.

Fixation: This representation aims to illustrate points of focus by displaying a single, unmoving fixation point as a red dot (Figure 4.7d). When a player maintains their fixation, the dot remains

[10] https://www.tobiigaming.com
[11] https://tobiicloud.com/share/

small in size. If the distance to the next fixation is within a threshold, the dot becomes larger. In real-time, this representation gives the appearance of small eruptions occurring on the interface that indicates whether a player is sustaining attention in an area (large dots) or flicking around between multiple areas (small dots). This representation was inspired by the 'Live Viewer' feature in the Tobii software, which Higuch et al. implemented for use in a remote collaborative scenario [104]. Our adaptation removes the elements of direction (tail, fade) used in other implementations to potentially reduce the mental load. Therefore, participants see a single point that appears then disappears instead of following a point.

Scanpath: The scanpath representation shows the length, duration and direction of gaze by using a sequence of fixations connected by a line (Figure 4.7e). Variation of this representation has long been used to visualise gaze patterns in both psychology, eye tracking usability studies and in HCI. This representation has been used unidirectionally such as in Qvarfordt and Zhai's study to show the interest of a remote unaware user, or bi-directionally such as in Jermann and Nussli to improve communication in a remote pair-programming scenario [119, 202]. We implemented this representation based on Goldberg and Helfman [92]. Many approaches have been presented to overcome the problem of visual clutter such as by bundling close fixations into a single point. Another approach is to use time duration such as in previous works (e.g. 5 seconds [250]. Instead, we opted to limit the number of fixation bundles displayed to 5, reduce the visibility of the lines connected the fixation bundle and removed other elements such the numbering of each fixation bundle to reduce visual information.

Fixation Trail: This representation was originally proposed by Li et al. to illustrate gaze paths as opposed to a single fixation point [153]. In our study, we renamed this representation to *Fixation Trail* (originally *Gaze Trail*) as it uses fixation data and leaves a trail of fixations (see Figure 4.7f). Li et al. believe that the availability of prolonged gaze information allowed participants to make references to multiple locations (e.g. contrasting two different fixation clusters). This representation is different to *Scanpath* in that we fixed the size of the fixation to a 30-px radius. Only 5 fixation points remain fully visible while the remainder fades over time.

AOI-Based Representations

In this category, we selected gaze representations that allow an observer to view the Areas of Interest (AOI) at which the user has looked as a whole rather than a specific point. Representations in this category use fixations aggregated over time. We can control how much visual information is displayed by adjusting its time window. Our implementation uses 120 fixation points ($\tilde{2}$ seconds) so as not occlude the screen (increase visibility, reduce visual information). We selected three representations with varying levels of visual information based on their popularity in eye tracking research.

Heatmap: Heatmaps represent the distribution of gaze points by a colour gradient (Figure 4.7g). We included this representation as it has commonly been used to visualise data in many domains and is not limited to eye tracking. In eye tracking usability studies it is also used for aggregating the gaze data from multiple participants. While the representation requires the aggregate participant data, it is possible to display a dynamic heatmap of a single tracked user [73]. Our implementation calculates the density of fixations within a 2-second window and displays the heatmap accordingly. The window size in our implementation attempts to balance visual information, allowing several 'hot spots' to be shown but do not remain for too long that it starts to occlude the interface.

Convex Hull: This representation is also known as *Cluster* (Tobii Pro) which display areas with a high concentration of gaze points as polygons to shows areas of interest (Figure 4.7h). Our implementation uses the fixation data stream in which we use to render a convex hull polygon in real-time using a clustering algorithm. As the stream continuously updates the convex hull, this creates a shifting polygon that shows the area in which the area is interested in. When the player looks rapidly between two points, the polygon stretches towards the points, but when the focus on a small area is high, the representation becomes small. This representation aims to provide minimal visual information of the area in which the participant has shown interest in.

Bee Swarm: Bee Swarm has been typically used to replay gaze points of several subjects against dynamic content such as videos, allowing efficient comparison of recordings from several subjects. Our variation aims to emulate this representation for use with a single user by displaying all fixation points in the last 2 seconds in which we enclose in a convex hull (Figure 4.7i). Each point was 5 pixels in radius with a red translucent fill. This visualisation allows the player not only

to see where the area in which the opponent was looking but the also points within that area. When the user focuses in a particular area, the points will become denser in the area, revealing an interest in the area. This representation has the highest amount of visual information as it combines characteristics from both AOI- and point-based representations.

Research Design

In this study, we aim to understand the following research questions derived from our review of the literature:

[RQ1] How well are human subjects able to infer intent from a range of gaze representations?

[RQ2] What characteristics of gaze representations enable humans to perform better in predicting intentions?

[RQ3] Do gaze representations have an influence on predicting intent in a competitive gameplay setting?

To answer these questions, we adopted a spectator ('third person') approach. This means that participants other than the players watched clips of gameplay sessions after the fact, with different clips featuring different gaze representations. We chose to make our participants spectators rather than active players in order to simplify the task they were asked to perform, as they could focus on interpreting the gaze data without worrying about managing their own strategy. This decision was informed by the common finding in prior works that live gaze data can be highly distracting from a task at hand [153, 189, 263]. As this was the first study to assess people's ability to infer intent from gaze data of a remote opponent in a game setting, we decided to begin with a more concentrated test of interpretive ability, before introducing the distractions of live gameplay in a future study.

We recorded our data in two stages. In the first, we recruited 8 players to record 4 games of our selected game—*Ticket to Ride* (described below) in which one player's gaze was shared. In the second, a separate group of 27 participants watched short selected clips (< 1 minute) of the recorded gameplay from the perspective of the player who could see the real-time gaze representations. These participants provided an assessment of each clip, which we used in our analysis. A secondary benefit of this two-stage approach was that we were able to conduct a

larger number of gaze visualisation assessments in a shorter period of time than would have been possible if all participants had had to actively complete multiple games. This enabled a larger sample size for this foundational work, which we will build on in future studies involving active players.

Application

We developed a custom networked system that overlays real-time gaze visualisations over any application (see through, click through). We used this application to stream gaze data between two laptops, each equipped with a Tobii EyeX eye tracker, in different rooms. We streamed two types of gaze data concurrently from the SDK: the lightly filtered gaze data and the sensitive fixation data. Our application is capable of displaying the gaze of one or both users simultaneously, meaning that we have the ability to switch on the gaze of one user to be visualised by themselves or by another user. Moreover, we can easily switch the gaze between representations through a control window on an extended screen.

Game and Stages

In *Ticket to Ride*, each player competes to claim train routes between cities across a map of North America, seeking to connect certain pairs of cities ('tickets') in a continuous path of linked routes. Only one player can claim each route, and players do not know which cities their opponent is seeking to connect. Therefore, players must place their routes carefully, staying alert to the risk that their opponent will claim a route before them and block the most convenient path between their goal cities. Keeping information hidden plays a large role in *Ticket to Ride*, as a player can gain significant advantages over their opponents by correctly guessing which routes they are planning to take. Hence, *Ticket to Ride* works as an excellent use case for our study for several reasons. First, it is a widely-played board game, already familiar to many, with a digital game version that is almost identical to the board game; and because it has relatively simple rules involving clear goals and actions, which can be learned in under 15 minutes by interested players who have not played it previously. Second, as a tabletop game, it demands constant attention, requiring regular thought processing and frequent changes to short- and long-term strategies. This affords us to obtain a range of behaviours to display. Third, Lankes et al. have determined

that turn-based games are better-suited to gaze transfer applications (as opposed to real-time first person shooters) [147]. This is because such games give players the opportunity to alternate between playing their own moves and watching their opponents' behaviour as they take their turn. Fourth, it is a map-based game, in which the players' focus and strategic calculations are closely concentrated on the geography of the map; gaze awareness and eye movement studies with map-based applications have been shown to be highly successful, as the static map gives users a point of reference [135, 202, 263].

As a playthrough of *Ticket to Ride* progresses, the map becomes increasingly populated with trains, and the strategic information available to the players becomes consequently richer while their choices of unclaimed routes become more constrained. Each player starts with a stock of 45 trains, and the game ends when any player's stock of unplaced trains falls below three. We use this train stock as a metric for game progression to divide the recorded games into three stages: early (45-31 trains), middle (31-16 trains), late (15-0 trains). The early stage is typically given over to planning, and placing trains are minimised as players try not to give away their goals. The middle stage sees greater competition for routes, and is where players often find themselves clashing over the most efficient paths between their respective goal cities. The late stage involves players either slotting in the last remaining routes that they need, seeking to block their opponent, or looking for other routes they can pick up extra points. We include an even selection of representations of each stage in our study, as we hypothesised that the usefulness and relevance of gaze visualisations might change as the game progresses.

Gaze Visualisation Recording

We recruited 8 players (4 pairs) to record the necessary gameplay clips between a *naive player* (whose gaze was being recorded) and an *aware player* (for whom their opponent's gaze was visualised). Both players were placed in separate rooms with an experimenter and given the opportunity to play the in-game tutorial for up to 10 minutes to understand the rules and mechanics of the game. Once the *naive player* was confident with the game, they were asked to perform the default calibration provided by the eye tracker. The *naive player* does not know that they are being watched to record and observe 'natural gaze'. Meanwhile, the *aware player* was briefed that different gaze visualisations would be shown to them and that their opponent

was not aware that their gaze was being shown. Each pair played three games, and the gaze representation was changed at every stage of the game. We recorded the displays using *Open Broadcaster Software (OBS) Studio*[12] and logged the gaze data with our custom application.

Following the recordings, we reviewed the videos and noted instances in which the naive player exhibited various behaviours, such as planning, checking game counters or tickets, and looking for routes to claim on the map. From the videos, we selected three short clips for each of the nine gaze visualisation formats: one in early game, one in middle game and one in late game. This gave us a total of 27 clips, exhibiting a variety of behaviours, and balanced between stages to control for the influence of game stage. Moreover, as different visualisations were used in different stages of different games, we are able to randomise the clips further. The length of the clips ranged between 20 and 60 seconds. All clips were saved without audio, to remove any potential sound clues such as the recorded player speaking aloud about their opponents' plans. As expected, the *aware players* experienced a level of distraction causing them to be drawn away from their own strategy and that some visualisations were preferred over others. However, we disregarded the feedback to avoid any early speculations.

Study

For the second stage of our study, we recruited 27 participants from the University of Melbourne. There were 11 male and 16 female participants, aged 18 to 44 years (mean=25.6), and included undergraduate students, graduate students and non-faculty staff. Only two had previously played the board game version of *Ticket to Ride*, but 17 indicated they had played a game of the same type in the past week. Participants were compensated with a $5 coffee voucher for their time. We encouraged participants to watch a short video[13] that describes the rules and objectives as part of the recruitment process.

Procedure

Upon arrival, participants were greeted and seated comfortably in a lounge room setting, facing a 48-inch (1080p) television display connected to a PC as depicted in Figure 4.8. They completed

[12]https://obsproject.com/download
[13]https://www.youtube.com/watch?v=uDODCHcj6Sg

Figure 4.8: Participant assessing a clip in a lounge room setting.

a consent form and a simple demographic questionnaire and were briefed on the rules of *Ticket to Ride*. The facilitator showed them an example clip of gameplay footage and with it, explained how they should assess each clip on a paper survey, with the measures outlined under 'Measures and Analysis'. Here, the facilitator ensured that the participants understood the rules of the game and what was required of them before starting the study. Each session ran between 30 to 60 minutes. We employed a repeated-measures design: each participant was shown all 27 (3 stages $\times$ 9 visualisations) clips, each of which was considered one study 'condition', representing one gaze visualisation format and one game stage. Participants were divided into three sets (A, B and C), and the order of the clips was randomised for each set.

Participants were given a maximum of 10 seconds to analyse the game and board state before the clip was played. In almost all cases, participants indicated that this was enough time for them to understand the game state, either with a verbal acknowledgement or a non-verbal indication such as a nod of the head. This time limit was imposed to prevent the participant from basing their inferences too heavily on reasoning about the layout of the game board, rather than the visualisations of the gaze data, and in order to more closely approximate a live gameplay setting in which the gaze information would be processed and responded to quickly. The facilitator encouraged participants to *think aloud* about what they were seeing, and to ask for any clarifications about the game if required. After the participant had viewed and assessed all 27 clips, they were finally asked to rank the nine gaze visualisation formats from most preferred

to least preferred. In summary, we collected a total of 27 participants $\times$ 9 visualisations $\times$ 3 stages $=$ 729 responses for each of our dependent measures.

Measures and Analysis

In response to each clip, participants were asked to record the following five pieces of information: (1) *Area of Interest Estimation*, using a map of the game board they circled the area (or multiple areas) in which they believed the gaze-tracked player had been interested. (2) *Route Prediction*, on the same map they highlighted any specific routes that they thought the gaze-tracked player was planning to build upon. They were also encouraged to write a short interpretation of what the gaze-tracked player was thinking during the clip or to explain their interpretation of the gaze behaviour out. (3) They rated how informative they found the gaze visualisation in the clip on a 7-point scale ranging from *Not informative at all* to *Extremely informative*. (4) They rated how well they felt they could predict the gaze-tracked player's intentions in the clip on a 10-point rating scale. Participants were permitted to skip any of these questions when they felt unable to give a response, for example, if they could not predict any routes upon which the gaze-tracked player was considering to build. Finally, we collected (5) *Qualitative Responses* at the end of the session, participants were asked to rank the representations in order of preference and explain the reasoning behind their choices.

Area of Interest Estimation and Route Prediction

To determine how closely each participant guessed the area of interest of the naive player, we compared the intersection of each response against the ground truth for the respective condition. Participants recorded their responses by circling the corresponding area in the map. To be able to analyse this data digitally, we scanned the highlighted pages and extracted the contours of the highlights using a custom-built OpenCV computer vision application. To establish a ground truth, we generated 2D density maps from the gaze data in each clip using the R statistical package. We then refined the AOIs against the manual analysis of the clips, discarding outliers that did not correspond to the routes the player eventually built. We then compared the intersection of the two regions, computing the precision (area of intersection divided by the total area circled by the participant) and the recall (area of the intersection divided by the area of the ground truth contour) of the participant's estimate (see Figure 4.9, top path).

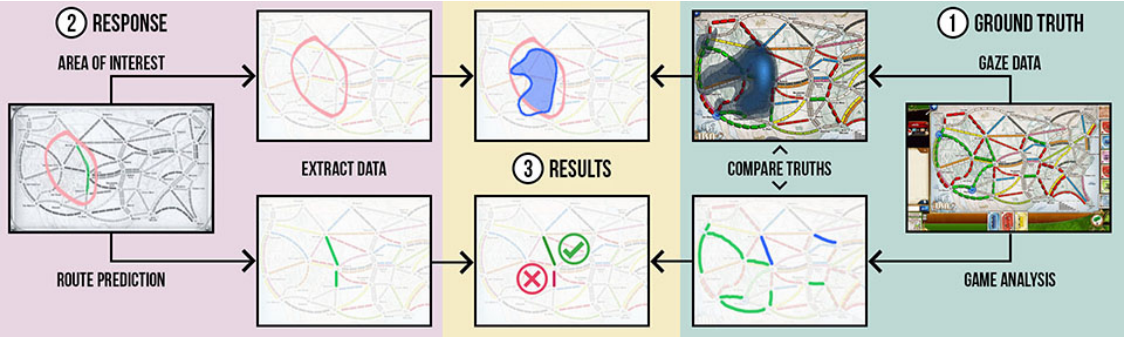


Figure 4.9: Method for analysis for Area of Interest (AOI) Estimation and Route Prediction.

Similarly to the analysis of the AOI Estimation, we scanned the highlighted maps and extracted the highlighted routes using our OpenCV application (see Figure 4.9, bottom path). We established the ground truth by recording the routes that the player eventually built after the clip. We also recorded routes that the player was considering to build but never did. We inferred such routes based on evidence from their hidden goal cards, subsequent game actions, and in-game comments. As with the AOI Estimation, we operationalise their performance primarily based on precision and recall, but also considered their harmonic mean i.e. *F-Score* ($F = 2 \times (precision \times recall) / (precision + recall)$). The details of these performance metrics are explained by Rijsbergen [209].

Quantitative and Qualitative Responses

For our third and fourth measure, we will use descriptive statistics to analyse how informative as well as the predictability of the gaze representation. To gain richer insights beyond the quantitative performance data, we collected subjective perceptions in three ways. First, we transcribed all verbal comments as they thought out loud when watching the clips. Second, in the questionnaires for the clips, participants answered an open-ended question that asked them to predict what the player was thinking. Third, we transcribed the interview about their order of preference at the end of the study. We were particularly interested in comments beyond the areas of interest and routes, including potential strategies, such as building a route to block the opponent, planning a route, verifying whether they have enough resources to build a certain route.

Results

We report the results of the nine visualisations at each of the three game stages and their overalls. Five primary metrics are reported: the percentage overlap of the AOI indicated by the participant against the AOI detected by the software; the performance between routes predicted by the participant and routes plausibly considered by the naive player; participant ratings of the visualisations' informativeness; participant self-assessments of how well they can predict what the naive player is planning to do; and participant rankings of gaze visualisation preference. Lastly, we compared the scores from their performance in our first two metrics against the last three.

Area of Interest Estimation

Representation	*Early*	*Mid*	*Late*	*Overall*
Heatmap	0.44 (0.78)	0.68 (0.63)	0.73 (0.75)	0.62 (0.72)
Fixation Trail	0.48 (0.41)	0.66 (0.72)	0.48 (0.75)	0.54 (0.62)
Dot	0.59 (0.49)	0.37 (0.55)	0.67 (0.36)	0.54 (0.47)
Cursor	0.46 (0.73)	0.68 (0.68)	0.44 (0.28)	0.53 (0.56)
Fixation	0.63 (0.72)	0.51 (0.87)	0.41 (0.80)	0.52 (0.80)
Scanpath	0.40(0.65)	0.50(0.79)	0.65 (0.70)	0.52 (0.71)
Bee Swarm	0.59(0.58)	0.36 (0.63)	0.43 (0.72)	0.46 (0.64)
Convex Hull	0.40(0.62)	(0.17 (0.44)	0.23 (0.44)	0.27 (0.50)
Spotlight	0.24 (0.37)	0.25 (0.43)	0.28 (0.65)	0.26 (0.48)

Table 4.2: AOI Estimation Results. Precision (Recall).

Table 4.2 shows the precision and recall for each representation, ordered by precision. The order of the performance of the visualisations in terms of precision were the *Heatmap* (0.62), followed by the *Fixation Trail* (0.54), and finally the *Dot* (0.54). This means that these are the best representations for selecting areas of interest within the ground truth. In terms of recall, the *Fixation* (0.80) was the best performing, followed by the *Heatmap* (0.72) and the *Scanpath* (0.71). This indicates that these were the representations in which the selected area captured a larger area within the area of interest. If we combine both metrics using the *F-Score*, we find that the overall best representation was the *Heatmap* (0.66), followed by the *Fixation* (0.63) and the *Scanpath* (0.63). To test whether the stage of the game had an effect on the precision and recall, we conducted a repeated-measures ANOVA for each representation. Except for *Spotlight*, we

found that the stage had a significant effect on the stage on the precision and recall. Precision significantly increased in the *Heatmap* and *Fixation Trail* as the game progressed, and decreased in the remaining conditions.

Route Prediction

Representation	*Early*	*Mid*	*Late*	*Overall*
Bee Swarm	0.70 (0.37)	0.67 (0.43)	0.43 (0.28)	0.60 (0.36)
Cursor	0.69 (0.85)	0.54 (0.37)	0.53 (0.19)	0.59 (0.47)
Heatmap	0.57 (0.31)	0.63 (0.29)	0.50(0.63)	0.57 (0.41)
Scanpath	0.40 (0.46)	0.45 (0.26)	0.63 (0.25)	0.50 (0.32)
Spotlight	0.50 (0.22)	0.49 (0.26)	0.25 (0.14)	0.41 (0.20)
Fixation Trail	0.44 (0.14)	0.50 (0.17)	0.26 (0.26)	0.40 (0.19)
Dot	0.59 (0.37)	0.37 (0.14)	(0.10(.04)	0.35 (0.18)
Fixation	0.37 (0.44)	0.26 (0.14)	0.27 (0.21)	0.30 (0.26)
Convex Hull	0.25 (0.11)	0.27 (0.10)	0.74 (0.02)	0.20 (0.08)

Table 4.3: Route Prediction Results. Precision (Recall).

Table 4.3 shows the computed precision and recall results for each visualisation ordered by precision. The best performing visualisations when it comes to route prediction in terms of *F-Score* are *Cursor* (0.52), *Heatmap* (0.48), and *Bee Swarm* (0.45). *Convex Hull* was ranked significantly lower in both precision and recall (0.20 and .08 respectively) and this was primarily from the late stage precision score (.02). The other low score noted, also in the late stage, is the *Dot* representation recall score (.04). We tested the effects of the gameplay stage on both precision and recall for each representation using repeated-measures ANOVA. For precision, only the *Scanpath* representation showed significance ($p = 0.014$) in which a Bonferroni post hoc test showed a significant difference between the mid and late stage ($p = 0.0069$). Here, we note that the *Scanpath* representation is the only representation to have increase in precision from the mid stage ($M = 45.5$) to the late stage ($M = 63.4$). A majority of representations, 6 out of 9 (*Bee Swarm, Heatmap, Spotlight, Fixation Trail, Scanpath and Convex Hull*), remained precise for the early and mid stages but significantly becomes less precise in the late stage. Both *Fixation* and *Cursor* precision dropped significantly from early ($M = 37.3, M = 69.1$) to mid game ($M = 27.2, M = 54.0$) but remained consistent in the late game. The precision of the *Dot* representation dropped linearly from early to late stage. There was no significant effect on recall on the gameplay stages.

Ratings of Informativeness

Representation	*Early*	*Mid*	*Late*	*Overall*
Spotlight	5.1 (1.0)	5.1 (1.1)	4.8 (1.0)	5.0 (1.0)
Heatmap	4.8 (1.0)	5.0 (0.8)	4.9 (1.0)	4.9 (0.9)
Cursor	4.9 (1.3)	4.8 (1.4)	4.6 (1.3)	4.8 (1.3)
Dot	5.1 (1.6)	4.8 (1.1)	4.1 (1.7)	4.7 (1.5)
Fixation	4.7 (1.0)	4.5 (1.2)	4.6 (1.1)	4.6 (1.1)
Scanpath	4.7 (1.0)	4.5 (1.2)	4.8 (1.0)	4.6 (1.1)
Bee Swarm	4.6 (1.1)	4.6 (1.3)	4.4 (1.4)	4.5 (1.3)
Fixation Trail	4.4 (1.2)	4.7 (1.0)	4.3 (1.0)	4.5 (1.1)
Convex Hull	3.2 (1.2)	3.6 (1.4)	3.0 (1.5)	3.3 (1.4)

Table 4.4: Ratings of Informativeness Results 1 Not Informative at all - 7 Extremely Informative). Score (Standard Deviation)

The responses follows a normal distribution (*Pearson chi-square normality test,* $p < 0.05$), in which calculated the mean scores and standard deviation for each visualisation for each stage, and overalls (see Table 4.4). We test the effects of stage on their rating on informativeness using repeated-measures ANOVA and found a significance in the mid and late stage ($p = 0.0247$). We tabulated the results and ranked them by overall rating accordingly. We found that *Spotlight* (5.0), *Heatmap* (4.9) and *Cursor* (4.8) received the highest rating while *Convex Hull* was rated significantly lower (3.3). *Heatmap* had the lowest deviation (0.9) among the visualisations showing an agreement among participants on its informativeness.

Self-Assessment of Predictive Ability

Representation	*Early*	*Mid*	*Late*	*Overall*
Heatmap	6.3 (2.0)	6.7 (1.9)	7.1 (1.7)	6.7 (1.9)
Cursor	7.0 (1.7)	6.1 (2.0)	6.6 (2.0)	6.6 (1.9)
Scanpath	6.0 (1.5)	6.0 (2.2)	6.6 (1.6)	6.2 (1.8)
Fixation	6.9 (1.7)	5.4 (2.1)	6.3 (1.5)	6.2 (1.9)
Spotlight	6.2 (1.7)	6.5 (1.5)	5.6 (1.9)	6.1 (1.8)
Bee Swarm	5.9 (2.1)	6.2 (2.0)	5.8 (2.3)	6.0 (2.1)
Fixation Trail	6.1 (1.7)	6.2 (1.5)	5.6 (2.2)	6.0 (2.0)
Dot	6.9 (2.3)	5.7 (1.8)	3.8 (2.0)	5.5 (2.4)
Convex Hull	4.0 (1.8)	4.7 (2.0)	3.9 (2.1)	4.2 (2.0)

Table 4.5: Predictive Ability Results. Score (Standard Deviation)

We ran the same statistical tests used in the previous measure and found that the responses also

follows a normal distribution. We calculated the mean scores and standard deviation of each visualisation for each stage and overalls, and ranked them by overall rating (see Table 4.5). We proceeded to test the effects of stage on the responses using repeated-measures ANOVA and found significance in the early and late stage ($p = 0.331$). The results shows that *Heatmap* (6.7), *Cursor* (6.6) and *Scanpath* (6.2) received the highest rating while *Dot* (5.5) and *Convex Hull* (4.2) was rated significantly lower than the others.

Preference Rankings

Figure 4.10 shows the ranking data for each representation by all participants, showing that the *Convex Hull* was ranked the lowest ($M = 2.0$) while *Heatmap* was the highest ranked ($M = 6.7$). *Cursor* and *Scanpath* also was ranked highly. We further analysed the ranked data using the Friedman test in which we found a significance ($p < 0.05$). The post hoc test showed that *Convex Hull* representation was ranked significantly lower than all other representations. We ran the Friedman test once again but this time removing the *Convex Hull* representation from the test. We found that there were no significance differences between the other rest ($p = 0.07547$). We proceeded to analyse the interviews to gain further insights into these preferences.

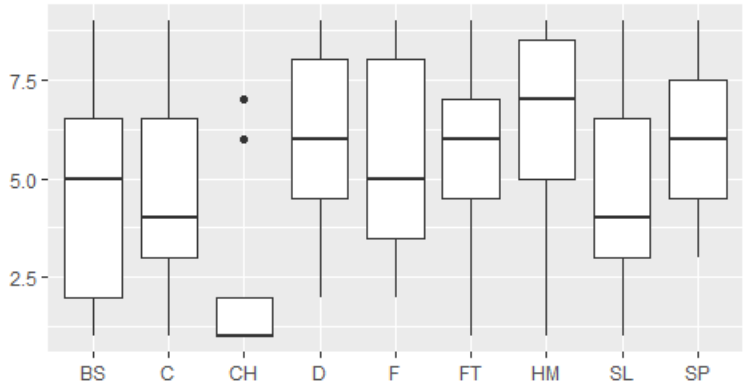


Figure 4.10: Boxplot of Preference Ratings given for each visualisation. (BS) Bee Swarm, (C) Circle, (CH) Convex Hull, (D) Dot, (F) Fixation, (FT) Fixation Trail, (HM) Heatmap, (SL) Spotlight and (SP) Scanpath.

We received many positive insights into why *Heatmap* was ranked the highest and preferred by a majority of participants. Participants that gave the *Heatmap* a high score mentioned its

'persistence' i.e. the visualisation gave them a moment to make an inference before it fades away (e.g. *"I like that it stayed on the screen for awhile" [P3]*). A small number of participants (e.g. P7) noted that they were familiar with the visualisation having seen or used them in other domains, noting its inherent usefulness. When participants were asked why they preferred *Heatmap* over the others, a majority they mentioned that they had difficulty following them. Two contributing factors for this were (1) speed (*"too quick" [P16]*, *"moved too much and too quickly"* [P11]) and (2) lack of conspicuousness (e.g. *"more subtle ones are harder to keep track of" [P19]*, *"it was hard to see" [P25]*). Some participants mentioned that the *Scanpath* visualisation blended into the game making it hard to differentiate (e.g. P20) while others found it useful (e.g. *"I like it because it gives you a track." [P11]*).

Conversely, some participants did rate visualisation that falls under point-based category highly. Participant 22 provided his explanation, stating that he preferred the point-based representations as he was able to process the visual information quickly and described trajectory- or AOI-based representations as "a lot going on". Participants often compared *Convex Hull* and *Bee Swarm* and found that the 'red dots' made it better (e.g. P17). *Bee Swarm*'s ability to have persistence yet show movement at the same time it is probably a reason why it scored well in the route prediction measure. Overall, participants provided rich feedback, noting that AOI-based representations were good for predicting overall strategy while point-based representations appear to be better moment-to-moment attention.

Discussion

Our study explored the characteristics of nine visual representations of gaze to determine which was most effective at conveying intent to an observer by exploring relationship between the characteristics of representations and human ability to infer intent from them. This section discusses the results from the five metrics as well as the post study interviews.

Representation Performance and Preference

Once again *Convex Hull* was largely indecipherable to participants, with the lowest scores for route prediction and close to the lowest scores for AOI estimation. However, the other AOI-based visualisations, *Bee Swarm* and *Heatmap*, had among the best results for route prediction. Unlike

Convex Hull, these visualisations can clearly display both the broad area of sustained interest and the focal point of the player's moment-to-moment attention. The *Heatmap* shows multiple hot spots and recent zones in which the player has looked. Interestingly, *Bee Swarm* performed well for route prediction but relatively poorly in AOI estimation measure. Our version of the *Bee Swarm* is encompassed by a convex hull display, which allows viewers to see at a glance the general region of the player's gaze while their more exact moment-to-moment attention is captured by a large number of rapidly-propagating red dots. We believe that participants used the convex hull to narrow their focus, enabling them to see the routes more clearly; avoiding a visual search for the next sequence.

Conversely, *Fixation Trail* and *Dot* performed strongly for AOI estimation, but relatively poor for route prediction, suggesting that these visualisations were effective at conveying general information but were not fine-grained enough for inferring specific intent. We believe that trajectory-based representations caused participants to divide their attention between two areas of interest and this was specifically mentioned by Participant 22. None of the point-based representations performed particularly well. We believe by the time participants start to make a connection between objects, the visual has moved somewhere else, causing them to start making a new inference elsewhere.

The *Heatmap* gaze representation overall scored highly in all five measures, and this is primarily attributed to its ability to provide a high level of visual information by using averaged information. Moreover, *Heatmap* takes a small amount of time to 'heat up' and cool down, allowing the player to know areas which are gaining interest and losing interest. We earlier noted that *Scanpath* performance increased from the early-mid to late stage of the game which may be reflective of the nature of the game. Towards the late stage players tend to track the missing paths which reduces the search space and enhances the performance of the *Scanpath* gaze representation. Further, with the increased availability of other cues to select from as the game progresses, a sustained gaze (presenting sustained interest) at a location is likely to signal with greater confidence the likelihood of an intention to action. To some degree, we expected the *Heatmap* to have performed the best being a well-used representation. On the other hand, the *Convex Hull* was notably disliked, being given the lowest or second-lowest ranking by all but four participants.

Preferences were mixed for the rest, with a wide variation in scores indicating that different

participants valued different characteristics. The trajectory-based representations were all ranked relatively highly; we hypothesise that this is because trajectory-based representations naturally calls attention to the connections between different points of interest as the player moves their gaze between them without overloading the user and therefore remained neutral in our results.

Interplay Between Game Cues and Overall Scores

As expected, we found that participants make use of in-game and behavioural cues to predict routes such as by observing what cards the naive player looked at or picked. From the game analysis, there was 17 showed cards being picked and many participants drew conclusions by matching the route seen and card colour which is a probable cause of differences in scores between the clips. Across all visualisations, prediction accuracy was lower in the late game than in the early and mid game. We consider several factors that may account for this trend, with different implications for the use of gaze visualisations in other contexts.

First, the visual clutter on the screen increases as the game progresses, so that in the late game there are substantially more train routes that may potentially distract the viewer from the gaze representation. Second, the number of available routes is lower in the late game, which tends to require players to draw more convoluted routes; the combination of zigzagging existing routes and fewer possibilities may increase the difficulty of recognising an obvious path that the player is aiming for. Third, the behaviour of the player is likely to be different, as they transition from long-term planning in the earlier stages to opportunistic route-grabbing and path-blocking in the late stage; this behaviour may be more difficult for a viewer to interpret. We believe that it is plausible that all three of these factors play a role, but based on our own observation of the game recordings, we expect that the third factor to be the most significant as gaze behaviour clearly shows a different pattern late in the game. This is exhibited in our results such as in the case of the *Dot* representation where the naive player was looking for was considering a route to build between a few potential routes.

Humans naturally rely on explicit actions as they confirm their thoughts, such as in real life. Regardless of representation shown, we are impressed at the ability of participants to put themselves into the mind of the player with the introduction of displayed gaze. In the same way, spectators rely on these actions, and therefore participants employ a variety of cues in

making inferences. However, we observed in our study that some participants better than other in keeping track of the elements in the game, allowing them to make better inferences.

Design Implications

Our results present various implications towards the digital games community. First and foremost, we demonstrated the possibilities of using different types of representations to display mental processes (e.g. intention) of players in real-time, which has direct implications towards *game developers*. When using gaze explicitly for use in games, game developers must be wary of how it is actually used in its execution. For example, when we the tested the *Convex Hull* representation, it worked well allowing two developers to reveal what the other was looking at.

However, during our evaluation, *Convex Hull* obtained consistently low scores revealing a flaw in our implementation when converting a static visualisation to real-time. Then again, this consistently low score gave us a reference point and confidence in the validity of our data. The design of gaze representations should further consider the amount it reveals without giving it away as it would give a player a disadvantaged. Imagine playing a game and your opponent is constantly one step ahead of you.

For the evaluation of games in real-time with gaze visualisation (implicit use), representations such as the *Heatmap* can be used to reveal intentions of players from natural gaze behaviour, and therefore provide insights into whether players understand the task in hand or is making correct strategic decisions. Second, a range of selectable gaze visualisations can implemented in live game streaming platforms such as Twitch, where *audiences* to create sensory excitement and anticipation. For example, audiences can anticipate intentions of live players before the actual move, providing a new experience. Lastly, *players* themselves can benefit from visualisations the gaze of their opponent in the form of an in-game 'power up' to create new gameplay experiences, especially in similar competitive contexts.

Contributions, Limitations and Future Work

This work presents three main contributions. First, we provide an understanding of the role of gaze and its importance to game designers in choosing and designing the right visualisation.

Second, an assessment of how people interpret visual gaze visualisations, which has not been done before in this context. The findings also highlight the effect of the different techniques on the inferring ability. Third, we contribute the assessment of AOI-based representations in the context of gameplay. Limitations wise, we used clips in our study from the entire game session which might not be comprehensive to show the goals of the whole game but enough to determine short term intention. Moreover, as each visualisation was tested in only three different clips, it is possible that one outlier clip could affect the overall result for that visualisation; nevertheless, we have taken this into account in our reporting of the findings. We also acknowledged that we only used an example of a strategic game to demonstrate gaze. However, we believe our findings apply to similar games.

Our future work focuses on *players*, where we give one player the ability to predict intentions using a real-time *Heatmap*, and with room for deceptive or misinterpreted behaviours to happen. We have also considered testing the viability of our findings with a different game (e.g. on with a less cluttered board) to provide further insights into the generalisation of our results. Furthermore, our work has further implications in gaze-aware artificial intelligence, and we have begun putting together a planning-based AI that predicts players potential intentions based on the recorded games from this study.

Conclusion

This paper contributes the groundwork for determining the characteristics of gaze representations that allow human players to make inferences about other players' intentions and to derive insights into potential strategies. This has been achieved by measuring the performance of human in their ability to predict intentions using a range of visual gaze representations. Our results showed us that human subjects have a remarkable ability to infer intent with the introduction of gaze. We conclude that representations that show both characteristics of point- and AOI-based representations would work well in our setting. Our work ultimately contributes gaze input as a viable implicit modality in multiplayer competitive computer games and with broader implications towards the games community.

End of Published Paper

No matter how brilliant your mind or strategy, if you're playing a solo game, you'll always lose out to a team.

– Reid Hoffman

5

Study 2: Effect of Gaze Awareness on Human-Human Intention Recognition

5.1 Introduction

This chapter presents Study 2, which was published as a peer-reviewed conference paper:

Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. 2018. 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* (CHI '18). ACM, New York, NY, USA, Paper 261, 12 pages. DOI: https://doi.org/10.1145/3173574.3173835

The paper has been reformatted to the formatting style of the thesis as recommended by The University of Melbourne's *Thesis with Publication*[1] guidelines. The paper in its original published format is found in Appendix A.2.

In the previous chapter, we found that humans can make inferences about intentions given the ability to see the gaze of a tracked 'unaware player'. Study 1 also finds that certain characteristics of gaze visualisation enable humans to improve their capacity to perform gaze-based intention recognition. Out of visualisations evaluated, we found that using the *real-time heatmap* performed best and was most preferred due to its ability to display a small window of aggregated gaze and current attention. This characteristic allows an observer to draw relationships with minimal clutter and without needing to follow a fast-moving live gaze point continually.

In this study, we extend the findings of the previous study in the following ways:

1. By using human players instead of human spectators, in which players will make use of the *real-time heatmap* visualisation during live gameplay to make inferences about their opponent's intended plans while also devising and managing their own strategic plans.
2. By assessing the ability of human players to interpret and respond to opponents' gaze cues throughout a live game, with a particular interest in the ability to make predictions during the early stages game when little or no in-game actions have been taken.
3. By introducing a deception aspect where one player is aware that their gaze is being tracked, creating an opportunity for the player to deliberately manipulate their gaze behaviour to potentially mislead their opponents. The combination of using an imperfect information game and a remote competitive setting affords us to explore this novel aspect.

[1] https://gradresearch.unimelb.edu.au/preparing-my-thesis/thesis-with-publication

As shown in Figure 5.1 below, this study is directly informed by Study 1, where we obtained a suitable visualisation for use in this study. Moreover, with the data set from Study 1, we began the construction of the agent model with promising results. The data set obtained from this study with its multiple conditions and full-game data allowed us to evaluate the artificial agent's capacity to perform intention recognition thoroughly in the following chapter (Chapter 6).

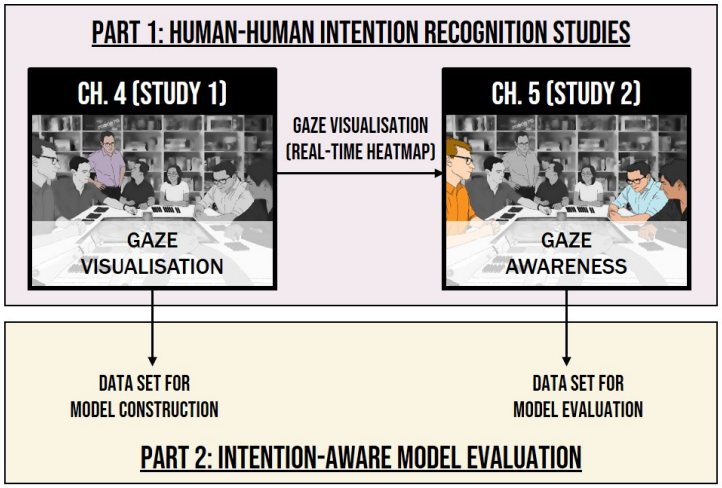


Figure 5.1: Part 1 contributes data sets to evaluate the artificial agent in Part 2.

In this study, we shift our focus from using *human spectators* to *human players* to study the effect of gaze awareness on intention recognition during live gameplay, forming sub-research question RQ1.2, along with further sub-research questions that helps us to answer RQ1 fully.

RQ1.2: What are the effects of gaze awareness for intention recognition in a competitive gameplay setting?

- **RQ1.2.1** Do human players with gaze visualisation perform intention recognition better than players without gaze visualisation in a competitive gameplay setting?
- **RQ1.2.2** Do human players with knowledge that their opponent is watching their gaze alter their gaze behaviour in a competitive gameplay setting?
- **RQ1.2.3** Can human players distinguish between non-deceptive and deceptive gaze behaviours in a competitive gameplay setting?
- **RQ1.2.4** Will human players perform intention recognition less accurately when the opponent knows that their gaze is being watched in a competitive gameplay setting?

In this study, we evaluate how well players can discern the overall strategic plans of their opponents when given the ability to see their gaze throughout the game using the *real-time heatmap* visualisation from the previous study. This study is examined through sub-research question RQ1.2 by comparing the results of three conditions using pairs of players: *no gaze visualisation*, *gaze visualisation of an unaware player*, and *gaze visualisation of an aware player*. We introduced the last condition to find out how players who are aware that their gaze is being tracked will manipulate their gaze to keep their true intentions hidden. In our pilot study, we tested a symmetrical setup where players can see a visualisation of each other's gaze on the interface, which showed that players constantly manipulate their gaze to keep their intentions hidden, but found that symmetrical conditions were not suited for our work. Hence, this study focuses on manipulating the two types of awareness asymmetrically instead: (1) *gaze awareness*, that is, when a player is aware of where their opponent is looking and (2) *awareness of being tracked*, that is, when a player gains awareness that their opponent is watching their gaze.

The overall results of the study showed that gaze awareness could further improve a player's capacity to recognise intentions of an opponent when displayed throughout the game, especially when compared to the baseline condition whereby he player can only make inferences based on the opponent's in-game actions. Further, the results showed that players could make early inferences solely using gaze visualisation and that players that were successful in keeping track of their opponent's intentions can potentially adjust their strategic plans accordingly.

However, despite the increased capacity to perform intention recognition, players also reported that they experienced a limited ability to keep track of their opponent's intentions and therefore unable to fully utilise the information that was theoretically available to them through the gaze visualisation. One possible explanation is that the opponent is continually assessing alternative plans, and the player needs to determine the likelihood of each as the game goes on and make alternate strategic decisions in response. Moreover, the player has to differentiate the priorities of the opponent, determine how the intentions relate to one another, and how each plan might affect their own strategic plans. Hence, successfully managing the multiple plans that are continuously being updated in a feedback loop and how they affect one another becomes a challenging task. Further, the study elicited similar reports on the *distracting* nature of using gaze visualisation when displayed over the interface as with previous studies (e.g. [65, 263]). The level of distraction

experienced may be partly attributed to the visualisation style, either due to its representation or the fact that it is a live visualisation, which suggests the potential for an alternative approach.

The study further finds that the capacity to perform intention recognition through gaze visualisation can be hampered when the opponent was made aware that their gaze was being watched. Players were successful in devising deception strategies that reduced their opponent's prediction accuracy about their plans, but also found degrees of difficulty in maintaining them (in line with [205]). All of the deception strategies we observed were conceived independently by the players, without suggestions from the researchers, and the stated inspiration behind many of the strategies was the player's own experience of observing the gaze visualisation in the previous round, which resulted in a novel *taxonomy of gaze-based deception strategies* (see Section 5.2).

To sum up, this chapter aims to find out the effects of gaze awareness for intention recognition in our chosen setting. The study adopts the *real-time heatmap* visualisation from the previous study, where it was the most accurately interpreted and second-most preferred visualisation style among the formats compared. The overall findings in this study indicated that although gaze awareness improves the capacity of human players to perform intention recognition in a competitive gameplay setting—answering RQ1, it also indicated that the task of performing intention recognition using the combination of gaze and in-game actions while keeping track of the state of the game is a cognitively demanding task. This demand is further increased as the human players have to balance other tasks simultaneously, such as managing their own strategy.

The first step towards our aim in this thesis is to provide artificial agents with the ability to perform gaze-based intention recognition, as we anticipated that it would narrow the interaction gap between the counterparts. Through this study, we collected a realistic data set that enables us to evaluate an artificial agent that incorporates gaze into its planning process in the next study (Chapter 6). Moreover, this study provides the motivations for delegating the task of gaze-based intention recognition to the agent, as the study finds that it is a cognitively demanding task for human subjects to perform. At the same time, delegating the task to the agent enables the player to move away from using gaze visualisation to infer intentions, which this study finds to have impacted the experience of players negatively. Therefore, by delegating the task of gaze-based intention recognition to the agent, it not only potentially reduces the confusing and distracting nature of using gaze visualisation, but it also enables us to address its inherent issue of *scalability*.

5.2 Published Paper: *Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay*

Start of Published Paper

Abstract

In competitive co-located gameplay, players use their opponents' gaze to make predictions about their plans while simultaneously managing their own gaze to avoid giving away their plans. This socially competitive dimension is lacking in most online games, where players are out of sight of each other. We conducted a lab study using a strategic online game; finding that (1) players are better at discerning their opponent's plans when shown a live visualisation of the opponent's gaze, and (2) players who are aware that their gaze is tracked will manipulate their gaze to keep their intentions hidden. We describe the strategies that players employed, to various degrees of success, to deceive their opponent through their gaze behaviour. This gaze-based deception adds an effortful and challenging aspect to the competition. Lastly, we discuss the various implications of our findings and its applicability for future game design.

Introduction

In face-to-face communication, our eyes have a dual function: to perceive the environment and to signal our point of interest within that environment [91]. In this context, gaze is an important behavioural cue that reveals our unspoken intentions to other people [49]; researchers call this "nonverbal leakage" [77]. As a result, interpreting gaze is a competitive element in many games, where players can gain advance warning of what their opponent intends to do by accurately reading their gaze. Variations on this tactic abound in all types of games, from poker players wearing sunglasses indoors to hide their eyes, to footballers using a "no look pass" to mislead opponents about the direction in which they will move the ball. As these examples show, counter-strategies for gaze interpretation range from the defensive (obscuring gaze) to the delusive (portraying a false intention through gaze).

By contrast, digital games rarely allow for gaze interpretation, as players are hidden behind screens and sometimes physically remote from each other. In a recent survey of gaze interaction in games, Velloso and Carter found an increasing number of applications in single-player scenarios [243], but limited usage in multiplayer and online games. HCI researchers have begun to explore the uses of shared gaze awareness in multiplayer games, finding that gaze visualisation can increase the feeling of social presence between remote players [147, 168]. However, despite its prevalence in face-to-face games, *gaze-based deception* in online gameplay has not yet been the subject of published research. This paper contributes the first study (to our knowledge) of human-to-human deception through gaze visualisation in an online strategic game.

In this study, we observed pairs of players engaged in an online turn-based strategy game with incomplete information, in the form of hidden objectives. Each player's gaze was tracked and, in some conditions, transmitted to their opponent and displayed as a dynamic heatmap overlaid on the game. The study involved three rounds of the game, corresponding to three study conditions: *no gaze visualisation*, *gaze visualisation of an unaware player*, and *gaze visualisation of an aware player*. First, consistent with the findings of our prior study [188], the findings of this study showed that players made more comprehensive and more precise predictions about their opponent's intentions when they were shown a gaze visualisation of an opponent that was unaware they were being tracked. Second, we find that players who knew their own gaze was being visualised were able to deceive their opponent with gaze, through both obscuring their true intentions and conveying false intentions. Following, we discuss the implications and applicability of these findings for future game design. Lastly, this paper contributes a typology of such gaze deception behaviours, categorised as *dissimulation* and *simulation deception strategies* according to the general theory of deception [254, 28].

Literature Review

Gaze and Intention

Gaze is a powerful form of nonverbal communication, as its direction alone can convey a number of signals, including one's current point of attention and future intentions [133]. Gaze often gives away this private information without our knowledge or intent, as "nonverbal leakage" [77]. It is common therefore for players in face-to-face games to peek at their opponents' eyes

to obtain clues as to what they will do next. Psychologists theorise that this inherent human ability to discern others' intentions stems from our *Theory of Mind*—the ability to put ourselves into another person's position and imagine how they might think and act—and that gaze is an important component to this ability [24, 151].

Gaze observation is a missing component in much technology-mediated human-human interaction. With the emergence of affordable eye trackers, researchers have begun to explore its uses in remote collaborative settings using visual representations of gaze (e.g. [11, 63, 153, 205]), and have shown that it can improve both communication and coordination. Similarly, a small body of work has emerged in recent years that explore eye-tracking technology as an input for digital games. Primarily this has concerned the creation of novel gaze-based gameplay experiences [240, 243], with only a few studies to date looking at shared eye tracking in competitive multiplayer games including our own [147, 188, 189, 205]. Typically, players of online games cannot see their opponents and so do not have access to the kinds of immediate behavioural cues (such as gaze) that enrich the experience of playing games in a physically shared location. In our previous work on the subject, we found that adding a visualisation of an unseen player's gaze into a digital game allowed spectators to make accurate predictions of what the player was planning to do, based on the areas of the screen receiving the most visual attention [188].

However, digital gaze visualisation is not a perfect replacement for face-to-face gaze observation. Researchers have found that live gaze visualisation from an eye tracker can be distracting and difficult for the observer to interpret, especially when there is a mismatch between gaze and other forms of communication [66, 153]. Furthermore, HCI researchers have found that the intended meaning of gaze behaviour can be difficult to discern through a visualisation [117], as it is often "hard to distinguish between deliberate communication and unwanted inputs" [168].

Gaze and Deception

As we use gaze cues to predict the intentions of others, gaze can be used against us as a method of deception. Players in many games mislead their opponents by simulating a fake intention with their gaze, as in a footballer's "no look pass". In *LaserViz*, the researchers simulated this mechanic using a two player shell game variation using cards, with one player wearing a head-mounted laser (that corresponds to the wearer's gaze) acting as the mixer [205]. The wearer initially

experienced a disadvantage as the opponent could correctly guess their plans based on the gaze visualisation but in subsequent rounds, the advantage was reversed as the wearer learned to trick their opponent by consciously looking at the wrong card.

The effectiveness of misleading gaze behaviour is predicted by Levine's *Truth-Default Theory* [152], which holds that humans have a strong tendency to presume that another person is communicating honestly, and that lies are usually only detected after-the-fact based on contradictory evidence rather than through observation of the liar's demeanour. Levine explains this as an adaptive tendency for most situations, as lies are much rarer than truths, but one that leaves humans vulnerable to deceit. We believe this theory can be generalised to scenarios in which the communication is solely nonverbal.

Studies have found that humans have a limited ability to differentiate between true and deceptive gaze behaviours (e.g. [83]) despite developing this ability at young age. Freire et al. found that children as young as four years old could detect small "leaked" gaze cues, and use them to infer that a person's private thoughts were contradictory to their verbal statements [84]. Then again, people are also able to adopt honest-seeming gaze behaviour when they lie to avoid suspicion. For example, Mann et al. found that people maintained eye contact when lying to foster trust, by taking advantage of the popular belief that liars avoid eye contact [165]. This suggests that distinguishing between true and deceptive gaze signals is a difficult task that could provide a rich resource for challenging social play.

Gaze and Plan Recognition

The problem of plan recognition has long been investigated by artificial intelligence researchers. Kautz and Allen define plan recognition as "the task of inferring intentions (in terms of plans) from the actions or utterances of observed actors" [130]. Plan recognition is an important aspect in real-time strategy games, as human players who discern the plans of their opponents can gain a significant advantage. Further, players with knowledge of their opponents plans will allow them to employ some form of deceptive behaviour. Cohen et al. outline two types of plan recognition, *keyhole* and *intended* [61]. *Keyhole recognition* is the inference of an agent's goals and plans through unobtrusive observation, as if through a keyhole. The agent has no intent to communicate its goals to the observer. *Intended recognition* is the conveyance of an intentionally

constructed impression of an agent's goals to the observer; it is typically found in adversarial settings, such as warfare, in which deceptive strategies are employed to thwart the recognition of the agent's true plans [51].

Strategic games have commonly been used as a context for adversarial plan recognition research (e.g. *StarCraft* [123]). Likewise, they have also been used in automated deception detection studies (e.g. *Mafia* [67]), which is a challenging problem that requires the combination multimodal inputs to effectively detect [2]. Although a large body of literature exists in these research areas, our interest for the purposes of this paper lies in plan recognition by human subjects. We seek to test whether a player who follows a live gaze visualisation will obtain the same benefits as *gaze following*, enabling them to discern their opponent's intentions and plans better. As previously mentioned, players in online games lack access to social cues, limiting their ability to discern plans until sufficient explicit information is provided. Byom and Multu explain that are three mechanisms of *Theory of Mind*: (1) knowledge of shared contexts, (2) perception of social cues, and (3) interpretation of actions, all of which are crucial when attempting to discern future behaviours [48]. The use of an online strategic game with incomplete information combined with the introduction of gaze visualisation allows us to study this effect. Furthermore, such games typically require players to construct hypotheses plans (that we can measure), and at the same time invite treachery, trickery and deception [213]. Additionally, playing strategic games have been found to activate regions of the brain associated with *Theory of Mind* [87]. Our work lies at several crossroads, as we begin to explore the potential of gaze as a playful nonverbal avenue for deception while drawing on its roots in face-to-face human communication.

Aims

We aim to explore the uses of gaze visualisation in online multiplayer games by understanding its potential usage for intention-reading and deception. Gaze, being a key signal for *Theory of Mind*, can assist a player to accurately imagine the mind of their opponent and thus help them to plant false beliefs. We formed the following hypotheses about this activity, based on our review of the literature:

[H1] Players who have knowledge of their opponent's gaze are better at discerning their opponent's overall plans.

[H2] Players who are aware that their gaze is being watched will alter their gaze behaviour to deceive their opponent.
[H3] When a player knows their gaze is being watched and attempts to mislead their opponent through deceptive gaze behaviour, the opponent will not detect the deception.
[H4] Players will be less accurate at predicting the plans of their opponent when the opponent knows their gaze is being watched than when the opponent does not know their gaze is being watched.

We aim to build upon our previous work [188] in three ways:

(1) By using players instead of spectators, where players will make use of gaze visualisation of their opponent during live gameplay, balancing the tasks of devising their own strategic plans with making inferences about their opponent.
(2) By assessing the ability of players to interpret and respond to opponents' gaze cues over the course of a whole game, with a particular interest in the ability to make long-term predictions during the early game before in-game actions have been taken. Previously, we were only able to obtain short-term predictions from the short clips used.
(3) By introducing a deception aspect where one player is aware that their gaze is being visualised, to create an opportunity for the player to deliberately manipulate their gaze behaviour to potentially mislead their opponents.

Study

We selected the computer game *Ticket to Ride* following the success of its use in our prior work [188]. In this game, players compete to build train routes between adjacent cities across a map of North America, and gain points for building a connection between specific pairs of cities that they are assigned on "ticket" cards (goal cities e.g. *Vancouver - Montreal*), which are unknown to the opponent. Players lose points if they fail to complete ticket cards they hold, and can block each other's paths, as only one player can claim each train route. Therefore, players must plan their train networks carefully to minimise the risk that an opponent will block them by claiming important routes first. More detailed information on the rules of *Ticket to Ride* can be found on the game's website[2].

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

There are several factors that make *Ticket to Ride* a suitable game for this study, in addition to keeping continuity with our previous study [188]. As a turn-based game, it gives players time to observe their opponent without having to focus on their own actions concurrently. It contains hidden information in the form of secret goals. As a map-based game, it encourages players to look at locations that correspond with their plans—and provides opportunities to do so misleadingly. Finally, it has a short gameplay time (around 30 minutes or less for two players), and is sufficiently simple that players are able to learn the rules and develop an understanding of the general strategies within a few minutes. We structured our study to ensure both players were familiar enough with the game to make predictions about their opponent, and that Player A (P_A) would be familiar with the gaze visualisation format by the third round (see Figure 5.2) to allow them to develop ideas about how to use it against their CU-GA opponent.

Study Conditions

Each pair of players played three rounds of *Ticket to Ride* against each other, with each player being subjected to a condition (see Figure 5.2). Each condition consists of the players' awareness of the study condition (**CA**: Condition Aware, **CU**: Condition Unaware) and their awareness of their opponent's gaze (**GA**: Gaze Aware, **GU**: Gaze Unaware). Condition-Aware players were given all details of the study condition for that round, including what their opponent knew. Each players' awareness differed between rounds, as described below. One player was aware of the condition in every round (P_A), while the other was naive after the first round (P_B). The second and third round correspond to *keyhole* and *intended* plan recognition scenarios, respectively [61].

Condition 1 (Baseline) — No Gaze Visualisation of Players

The first round consisted of a standard online match with no gaze visualisation and both players are aware of this, making both players CA-GU (see Figure 5.2). This round served as a baseline for recording the players' behaviour under normal gameplay conditions. It also allowed players to get experience playing the game to minimise any learning effects, and to get a sense of forming their own strategy while discerning their opponent's strategy, based only on explicit in-game actions.

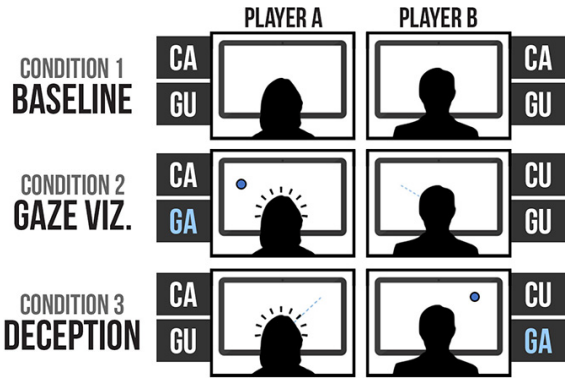


Figure 5.2: Study Conditions (Player A: P_A, Player B: P_B). In Condition 1, no gaze data is visible to either player. In Condition 2, P_A can see P_B's gaze, and P_B is unaware of this. In Condition 3, P_B can see P_A's gaze; P_A is aware of this, and that P_B does not know P_A is aware.

Condition 2 (Gaze Viz.) — Gaze Visualisation of Unaware Player

In the second round, P_A was shown the gaze of P_B throughout the game, and knew that P_B was not aware their gaze could be seen (see Figure 5.2). This allowed P_A, as a CA-GA player, to try to discern the intentions of an entirely naive opponent (CU-GU) through their gaze natural gaze behaviour.

Condition 3 (Deception) — Gaze Visualisation of Aware Player

In the third round, P_B received the gaze visualisation of P_A. But this time P_A knew their gaze was being tracked (see Figure 5.2). Not knowing this, P_B (CU-GA) attempted to discern the intentions of P_A (CA-GU) through their gaze behaviour without knowing that they could be intentionally misled.

Recruitment

We recruited 40 players (22 female) from the University of Melbourne, and assigned them into 20 pairs. These included undergraduate students, graduate students and non-faculty staff, aged between 18 and 39 years (M=23.2). Players were compensated with a $20 (AUD) gift card for their time. Before the study, players were asked to learn the rules of the game by watching a video tutorial, if necessary; 11 had played the game previously. A majority self-identified as an "occasional gamer" (23), while the rest self-identified as a "non-gamer" (11) or "regular gamer" (6). Ten players had used eye tracking previously, all as participants in a research setting.

Experimental Setup

The study was conducted in a university usability lab with two players in each session under approval by the University's ethics committee. To simulate online gameplay, we placed players in separate observation rooms and set up an online game of *Ticket to Ride* between them. We provided snacks and bottled water during the study in addition to the compensation due to the duration of each session. Both computers were fitted with a Tobii 4C eye tracker[3], and the gaze data from the eye tracker was streamed between the computers using a custom networked system that overlaid real-time gaze visualisations over any application (also used in our previous study [188]). Players were also given a tablet computer that mirrored their desktop computer screen and instructed to sketch notes about their opponent's strategy on the tablet using a stylus. We placed a video camera in each room to record the players, and the computer screens were also recorded for the duration of each session. Lastly, we logged the raw gaze data of both players for each game round for later analysis.

Procedure

Two researchers were present in each session to facilitate the study and to observe the players. The researchers followed a well-rehearsed script, so as not to reveal the study conditions to the players. On commencement, both players were given an initial briefing together that explained that their gaze would be tracked throughout the study for the sole purposes of later analysis. Players were then randomly allocated into one of the two observation rooms, with one researcher in each. There, players were given a written overview of the study, consent form and basic demographic questionnaire to fill out. Following this, players were calibrated using default calibration procedure of the eye tracking device, and instructed to play the game's interactive tutorial until both players were satisfied that they understood the game for up to 10 minutes.

Players then played three rounds of *Ticket to Ride* against each other through an online connection. To ensure timely completion, each player was given a 15 minute cumulative time allowance for their total turns in each round; if either player ran out of time, we manually calculated the scores for that round. At the start of each round, we requested each player to pick all three randomly assigned 'ticket' cards to complete (each representing a pair of 'goal cities', potentially

[3] http://tobiigaming.com/eye-tracker-4c/

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. The researchers prompted the player occasionally on these topics, and took notes throughout the session on player comments or behaviours related to the study objectives. Researchers were permitted to respond to any questions regarding the rules of the game at any time during the session. In the conditions where the player could see gaze, we provided a short tutorial on how we would like them to annotate on the tablet—*circling* the cities they thought were their opponent's goals and *highlighting* the routes they thought their opponent planned to take (see Figure 5.3). The gaze was visualised as a dynamic real-time heatmap, as our previous study showed that this was the best representation for discerning the goals of a player based solely on a gaze visualisation [188]. The dynamic heatmap visualisation was semi-transparent, and the application allowed the player to click "through" it as normal.

At the end of each round, players completed a brief section of the questionnaire (see *Measures and Analysis* section below for questionnaire details) about the round they just played. At the end of all three rounds, both players were invited to add any final thoughts, and the CU-GA player was asked whether they suspected any deception in the study. Finally, both players were fully debriefed about the study conditions.

Measures and Analysis

Player Self-Assessment and Experience

We created two sets of questionnaires for players to complete following each round of the game, with slightly different questions for each. In both sets, the first three questions following all conditions asked the player to rate their ability to predict their opponent's short-term and long-term plans, and what aspects of the game they had found helpful in predicting their opponent's plans. These repeated measures allowed an assessment of the subjective effects of gaze visualisation and deception on intention recognition ability. We expected that players would make a higher number of predictions when gaze was visualised and that these predictions would be more accurate when the opponent was unaware their gaze was being visualised. To test this effect, we ran a multiple regression model on the questionnaire responses, with independent variables being gaze awareness (whether the opponent's gaze was visualised on their screen

or not) and condition awareness (whether the opponent was aware that their gaze was being visualised or not).

For the second and third round, the questionnaire asked additional questions to assess the impacts of gaze visualisation. Both players were asked to rate their experience with the gaze visualisation following the round in which they saw it. After the third round, the CA-GU player was asked whether they attempted to deceive their opponent and how. To test whether the deception of the study had been maintained, the CU player was asked after the second round in what ways the game had been different to the first round; and after the third round, the penultimate question asked whether they felt they had been deceived about the study at any point.

Prediction Performance

We are primarily interested in players' actual success at discerning the plans of their opponents, as opposed to their self-assessed abilities. To determine this, we first extracted the ground truth (all routes each player claimed) for each individual game for each player (120 games total). We then analysed the study recordings to list all instances in which a player made an inference about their opponent's plan, including both verbal statements and tablet annotations. We then categorised these into *goal predictions* (goal cities the opponent is ultimately connecting to, i.e. their secret objectives) and *subgoal predictions* (what cities or routes they intend to connect through to reach their goal cities). We measured each of these predictions against the opponent's ground truth, defined as the secret objective cities they were assigned by the game, the routes they ultimately claimed, and any routes they verbally declared they were considering but did not end up claiming. We included the latter because players could change their strategy in response to events during the game, particularly after being blocked by an opponent who had correctly guessed their plan. We excluded from the ground truth some incidental routes, primarily any last-minute building decisions in the final turn that were clearly not part of their planning before that point in the game.

We assessed players' prediction performance using three metrics: (1) their accuracy in discerning goals, (2) the extent to which they discerned their opponent's plans, and (3) the *distance* of each guess from the ground truth, in terms of both goals and subgoals. We evaluated players' goal and plan prediction performance by employing the $F_1 - Score$ ($F_1 = 2 \times (precision \times$

$recall) / (precision + recall)$) metric from information retrieval literature [209]. We compared the number of goal cities that were guessed correctly against the total number of cities guessed (*Precision*), and against the number of cities that were in the ground truth (*Recall*). We then performed the same comparisons, replacing the number of correctly-guessed goal cities with the number of goal cities that had been guessed as subgoals. A $F_1 - Score$ value was calculated from both sets of comparisons. Lastly, we employed the *distance* metric from graph theory [252]. As the playing area of *Ticket to Ride* is laid out as a graph diagram, it is possible to determine an error distance by treating all cities as nodes in a graph, and measuring how many nodes the guess was away from the ground truth [157]. To simplify our analysis, we take the shortest distance between the guess to any goal (or subgoal) from the ground truth. Predictions that were adjacent to a goal (or subgoal) thus have a distance of 1, and predictions that correctly matched the ground truth have a distance of 0.

Results

Each session lasted for at least two hours, and consisted of three rounds of *Ticket to Ride* between two players, for a total of 60 rounds or 120 rounds if each player is considered separately. We then compared the extracted predictions made for each player in all conditions against the ground truth for that respective game. The use of the "think-aloud" protocol in our study provided rich descriptions of players' strategies and considerations throughout each of the games. These reflections gave us confidence in our analyses of the game data, as the results are in accordance with what players said themselves.

Effect of Gaze Awareness and Deception

Effect on Player's Self-Perceived Predictive Ability

We examine players' self-perceived predictive ability, i.e. how well they *thought* they discerned the plans of their opponent. We built a multiple regression model to predict player's self-ratings on this measure. A significant regression equation was found for both short-term (R^2*=0.23, F(3,116)=11.28, p<0.05*) and long-term (R^2*=0.20, F(3,116)=9.42, p<0.05*) predictive ability. We can define the model based on the latter as follows:

$$Response = 3.27 + (1.8\,(Gaze)) + (-1.32\,(Deception))$$

Both measures are significantly correlated across all conditions for both players (*Pearson(r)=0.78, p<0.05*), meaning that if a player rated their ability to predict specific plans highly, their ability to predict overall goals was also rated highly. The model also shows that there were no learning effects in predictive ability, but there was an increase in players' confidence in their ability to form strategies and make predictions.

Effect on Player's Outcomes and Experience

We first report on players' self-assessments (questionnaire responses), each measured on a 7-point Likert scale (1 being lowest agreement, 7 being highest agreement). Subsequently, we report on the way players spoke about the experience of the game, and the level of success that they achieved.

Overall, players found it easier to make short-term predictions (moves) (P_A*=5.6,* P_B*=5.4*) and long-term predictions (strategy) (P_A*=5.1,* P_B*=5.3*) in conditions where gaze was visualised. Both sets of players noted the gaze visualisation was useful for formulating their own strategy (P_A*=4.9,* P_B*=5.1*), and that it changed the way they played the game (P_A*=5.4,* P_B*=5.2*). Both sets of players also believed that gaze visualisation had an influence on the outcome of the game (P_A*=5.1,* P_B*=4.7*). In Condition 3, the CA-GU players indicated that they changed the way they played the game as they knew that their opponent could see where they were looking (*M=5.2, SD=1.6).* Consequently, the players were somewhat bothered by the knowledge that their gaze was being observed (*M=3.9, SD=1.7*), and all of them employed some form of deceptive strategy to counteract this (see *Gaze-Based Deception Strategies* section).

In a comparison of the actual game outcomes in Conditions 2 and 3, we found that the CA players were no more successful than the CU players. This was despite their information advantage, as demonstrated by their greater ability to discern the opponent's true intentions (see Table 5.1). To explain this, we refer to the comments from the CA player for both conditions. In Condition 2, CA-GA players commented that the gaze visualisation was distracting and drew their attention away from their own plans, both by providing additional information about their opponent's plans and by creating visual clutter on the screen. We believe that this was

				Goal Prediction		Plan Prediction		Distance	
Condition	*Player*	*Won*	*Avg. Predictions*	$F_1 - Score^1$	$F_1 - Score^2$	*Precision (Recall)*	$F_1 - Score$	*Total*	*Mean*
1 (Baseline)	P_A (CA-GU)	13	5.8	0.06	0.13	0.49 (0.13)	0.19	25	0.22
	P_B (CA-GU)	7	4.5	0.11	0.19	0.51 (0.11)	0.17	29	0.32
2 (Gaze Viz.)	P_A (CA-GA)	11	12.7	0.23	0.44	0.48 (0.34)	0.39	75	0.30
	P_B (CU-GU)	9	5.8	0.02	0.04	0.49 (0.13)	0.18	24	0.21
3 (Deception)	P_A (CA-GU)	12	4.0	0.09	0.17	0.48 (0.09)	0.14	24	0.30
	P_B (CU-GA)	8	11.7	0.18	0.35	0.41 (0.28)	0.31	111	0.48

Table 5.1: Prediction Performance Results. The F_1–$Score^2$ is calculated based on all cities (goals and subgoals) guessed against the ground truth to find out if players did discern the goal cities did not recognise them as goal cities while F_1–$Score^1$ is calculated based on only correctly guessed goal cities.

likely exacerbated by the fact that real-time gaze visualisations were an unfamiliar interface element for all players. The CA-GA players also commented that the visualisation influenced them towards a 'blocking' strategy, which was focused on interfering with their opponent rather than pursuing their own goals. This strategy proved relatively ineffective in many cases, as it caused the aggressor to waste as many turns and resources as the player they were trying to block. For example, $P6_A$ stopped their opponent from completing 2 out of 3 ticket cards, but completed none of their own objectives. The CA-GU players spoke about the difficulty of maintaining their deceptive strategies while also thinking about their plans, especially early in the game, noting that it was the best time to be deceptive as their opponent had no other cues. This difficulty in maintaining deception echoes the findings of the *LaserViz* study [205]. Many of the CA-GA players commented that they felt like they were 'cheating' in the second round with an 'unfair advantage', whereas the final round felt more challenging, and generally more exciting. Nevertheless, the CA-GU players found it interesting and engaging to know that their gaze was being watched, and stated that it gave the game added depth.

At the conclusion of each study session, the deceived player was asked whether they suspected that they had been deceived at any point. Only one player expressed some suspicion while the rest did not express any at all. All players expressed the opinion that both they and their opponent had improved their performance each round, which they attributed to ordinary factors such as the luck of the draw or taking greater care in planning their strategies.

Effect on Prediction Performance

Table 5.1 displays a summary of prediction performance results for all three conditions. A total of 889 predictions (guesses) were made by both players across all conditions. The introduction of gaze increased the average number of raw predictions made compared to when players could not see gaze. Further, we noted that players were able to make their first predictions significantly earlier when given the ability to visualise the gaze of their opponent than in the other conditions without gaze, on average within the first minute from the start of the game. In the conditions without gaze visualisation, the first prediction on the opponent's plans was only made after the opponent claimed a few routes, typically a few minutes into the game. Predictions did become increasingly difficult if an opponent was effective in hiding their intentions. On occasion, the

Figure 5.3: Ticket to Ride Gaze Visualisation Prediction Example. The three figures above shows three stages (early, mid, late) of the same game from left to right, and the real-time heatmap visualisation used in our study. The green tracks belong to the CA-GA player (who has annotations in yellow highlight) while the red belongs to CU-GU player. In early stage, the CA-GA has already predicted part of their opponent's plans, discerning that the opponent is interested in a particular route(s) when no trains have been played yet by the opponent. In the mid stage, we can see how the predicted strategy has been played out and now better discern a different part of their plan. In the late stage, we can see more of the predictions materialising. The figure has been enhanced by reducing the contrast of the background to better show the real-time visualisation, paths claimed and the annotations.

researchers would prompt the player on what they thought their opponent's plans might be, which would often result in a negative response from the player (e.g. *"I don't know", "I am not certain", etc.*), or a brief response with no solid prediction that the researcher could record.

A reason for low prediction counts, particularly in the conditions where players were gaze unaware, is that players often focused on their own plans initially and were less concerned about their opponent's plans until completing their own. This was more likely in round 1 as new players got used to the game. Players may have been particularly good at hiding their plans, in line with the game design, making it hard to make predictions. This led to the predicting players only making predictions when they were confident, leading to a low number of average predictions with high precision but low recall.

In the conditions where players were gaze aware, they explicitly mentioned that they used the gaze visualisation extensively to make predictions, especially in the early stage of the game. This was as expected, as few actions are played in the early stage and the formation of their own plans was still in its infancy. We highlight that players used a variety of in-game actions as well to make predictions, such as observing which colour cards the opponent picked from the deck and matching that to the colour of routes on the board, or observing the general direction in which the opponent was building. Experienced players used the average distance between pairs of cities to gauge whether their opponent had completed a ticket. Other in-game cues contributed to their predictive ability, such as observing which train cards were not selected and analysing

opponent's building patterns. Players who frequently annotated on the tablet found it effective for keeping track of their opponent's plans.

In the third condition, the CA-GU players made fewer predictions about their opponent's plans than in the previous rounds. We believe this is for three reasons. Firstly, players focused first on deceiving their opponent and second on attempting to complete their own goals, leaving little attention for discerning their opponent's plans (*"I wasn't focused on predicting their movements but rather on how to throw them off since I know I'm being watched."* [$P9_A$]). Secondly, the deceptive gaze activity required these players to keep track of two sets of plans (true and deceptive), making it hard to keep track of a third. Thirdly, players felt less confident about their predictive abilities after seeing the gaze visualisation and then having it removed (*"I was less able to assess my opponent's strategy than I thought I could. It allowed to see the real advantage of having the eye tracking in the second round. With the eye tracking, I would have been able to see if she was looking at the east coast the entire time."* [$P11_A$]).

Upon calculating the different metrics to determine the effect of gaze and deception on prediction performance, we found that the CA-GA players performed the best overall in Condition 2 amongst the other rounds, despite having the highest number of predictions. There was some degree of error indicated by the total distance (75), but this appears to be in the normal range when averaged, meaning that the predictions made did indeed have a distance but was of low value. In contrast, performance scores for the CU-GA players in Condition 3 reveal that they still made more correct predictions based on the goal and plan prediction scores compared to conditions with no gaze. Further, the distance metric did reveal that only 36.6% of the CU-GA players predictions had a distance of over 1 (CA-GA was 26%), showing that the players could still judge their opponents' plans with reasonable accuracy.

We expected the likelihood of players discerning the goals of their opponents to be generally low, especially in conditions without gaze, and this proved to be true ($F_1 - Score^1$). As a further step, we wanted to see if players would guess their opponents' goals as part of a route (i.e. as subgoals) without discerning them as end goals (i.e. without circling them or stating that they were a 'ticket' destination). The results for this analysis were higher ($F_1 - Score^2$), meaning that players did indeed discern goal cities without recognising them as end goals. A further comparison revealed that CU-GU players in Condition 2 had a constant low prediction score, which we

believe is attributable to three factors. First, the CA-GA players kept their plans hidden better, by playing a more reserved strategy. Second, the CA-GA players often employed a blocking strategy, causing the CU-GU players to make incorrect or fewer predictions. Third, the CA-GA players completed less of their own intended plans in this round, so there was less on-board information for the CU-GA players to make inferences from. As an additional point, we observed that players commonly extended their routes beyond their goal cities to get bonus points (*Ticket to Ride* gives bonus points to the player with the longest overall route), which made it harder to discern whether the 'ticket' goal was the end of a route or not. In Condition 3, CA-GU players (e.g. $P15_A$) commonly looked beyond their goal cities as a form of *gaze misdirection*.

The CA-GA players in Condition 2 scored an average plan prediction score of 0.39 (*SE=0.031*) when they could see their opponent's gaze, compared to 0.19 (*SE=0.036*) in Condition 1 when they could not. A Welch Two Sample t-test indicated a significant difference between rounds (*t=4.20, df=37.23, p<0.05)*, along with a medium-sized effect *r=0.57*. We thus reject the null hypothesis [H1]. The CU-GA players in Condition 3 scored an average distance error of 0.48 (*SE=0.050*), whereas the CA-GU players in Condition 2 scored an average distance error of 0.3 (*SE=0.033*). A Welch Two Sample t-test indicated a significant difference between rounds (*t=3.01, df=413.32, p<0.05*), with a small-sized effect *r=0.15*. We thus reject the null hypothesis [H4].

Summary

Our results showed that gaze added value when used to discern plans in a strategic game, especially early in the game, consequently revealing a large part of the tracked player's plans as the game progressed. Figure 5.3 illustrates how gaze was displayed, followed by its early and partial prediction of an opponent's plan. Further, we found that deception did have an effect on predictive performance, though the player's true intentions still leaked through their gaze. The CA-GU players did need to look at their real plans on occasion and, over time, gave them away as they claimed more routes. As a result, CU-GA players guessed both true and deceptive plans. The results are also consistent with Hypothesis [H3], as the CU-GA players did not state that they were being deceived (as predicted by the *Truth-Default Theory* [152]), when in fact all the CA-GU players did attempt to deceive. Gaze visualisation had overall positive effects on both player experience and player performance, although it was also found to be 'distracting'. The

findings also show that the deception condition did cause the CA-GU players to change their gaze behaviour, as they became more aware of where they were looking and sought to deceive their opponent, validating Hypothesis [H2]. The strategies employed are summarised in the section below.

Gaze-Based Deception Strategies

All CA-GU players in the final round stated that they attempted to deceive their opponent by manipulating their gaze behaviour, and described one or more strategies they used to do so. We categorised the strategies according to the general theory of deception, which divides deception methods into two main categories: *dissimulation* (covert, hiding what is real) and *simulation* (overt, showing the false) [254, 28]. Within these categories, we have created a more specific typology that draws on both the general theory and deception concepts in military strategy, such as camouflaging [74].

Dissimulation Deception Strategies

We observed several ways that players sought to conceal their plans when they knew their gaze was being observed. The most frequently employed and direct dissimulation strategies were *Gaze Averting* (6) and *Information Reduction* (6), followed by *Gaze Scattering* (5). The least employed strategies, *Risk reduction* (3) and *Obstruction* (4) were more oblique, seeking to deny the opponent an accurate target to uncover.

Gaze Averting

Players sought to avoid looking at the routes and cities that were of interest to them. In Bell and Whaley's typology, this was a kind of *masking* [28]. The practice was difficult to sustain; players frequently expressed frustration at the conflict between their desire to see what they were doing and their anticipation that doing so would give away their plans to the opponent. From observing and listening to players, it was apparent that this avoidance tactic took concentrated effort to maintain, and as such, it tended to lapse for increasing lengths of time as the game went on. In face-to-face communication, gaze averting is often attributed as a tell-tale sign of someone lying [76], but this was not detected in our study.

Gaze Scattering

Players went a step further than aversion, by actively scattering their gaze around the screen to give the impression that they were looking "everywhere". This tactic was intended to make it hard to distinguish moments of real planning from moments of arbitrary looking—similar to Bell and Whaley's concept of *dazzling* [28]. Gaze scattering also served as a kind of attack on the opponent's concentration. In Condition 2, all of the CA-GA players noted that the gaze visualisation distracted them from planning their own strategy under time pressure. In turn, players replicate this distraction for their opponent in Condition 3 by scattering their gaze, essentially "weaponising" the gaze visualisation.

Information Reduction

Players employed several tactics to reduce the amount of information their actions could give away. For example, players favoured drawing cards from the blind deck over drawing face-up cards, so that the opponent could not see what colours they were collecting. Other players took their turns as quickly as possible, minimising their opponent's time-frame for making inferences about their gaze and actions, causing the opponent to concentrate on their own plans instead.

Risk Reduction

Players adopted a more flexible strategy when they were aware that their gaze was being observed, to minimise the effect of being blocked by their opponent. For example, players chose to build across areas that offered the most alternative pathways, rather than going for a more direct and efficient route. Players who adopted this strategy typically tried not to claim any train routes for the first half of the game, and focused instead on collecting cards; this both reduced the information available to the opponent (gaze and in-game), and allowed the player to re-plan their route if the opponent's actions got in the way of their initial plan. Risk reduction strategies such as these ensured the player's intentions could not be pinned down, because they were left open to change.

Obstruction

Players sought to disrupt their opponent as a form of counter-attack, taking the focus away from their plans by playing aggressive moves such as taking cards and routes that they thought their opponent might need. The intended benefits of this were twofold. First, if done well, it forces the opponent to use up their resources to work around the blocking moves, which ideally left the opponent unable to complete their objectives by the game's end. Second, by forcing the opponent to play reactively, it neutralises the opponent's information advantage as the opponent would be less concerned about the gaze-tracked player's secret objectives. This was a high-risk strategy, as the player attempting it had less information available to them to know whether their aggressive moves were, in fact, obstructing their opponent's plans.

Simulation Deception Strategies

As well as hiding their true intentions, players sought to actively mislead their opponent through various kinds of *Gaze Misdirection* (3), a specific example of which was *Decoys* (16). Players also used *Gaze Camouflage* (4) strategies to create a false impression about the meaning of their actions.

Gaze Misdirection

Players attempted to give their opponent a false idea about what they were thinking by exhibiting behaviour that was consistent with an imagined alternative plan. For example, the player might stare at a route they think the opponent wants to build on, to scare them into thinking they will be blocked; this was intended to induce a degree of panic and to keep them from noticing the player's real plans. At times, players would take this further by alternating fixations between the route and a face-up card of the matching colour, to heighten the impression that the player was planning out their own construction, rather than idly looking at the map.

Decoys

A common sub-case of gaze misdirection was the use of decoy paths and is the most common deception strategy employed. Players who employed this strategy chose a set of cities far away from their real objectives, and repeatedly traced a path between them to give the impression that

they were planning to build there. The decoy was usually chosen at the beginning of the game, shortly after the player had received their real objectives. Players typically watched out for their opponent building on their decoy path, and took this as confirmation that they had deceived the opponent. For the disinformation to be effective, however, the enemy must know how it is being accepted and interpreted by the target [113].

This strategy appears to be the most effective, particularly as opponents showed a degree of confirmation bias in their observations. Typically, the opponent expressed an early suspicion about one or two locations that the CA-GU player had looked at, and subsequently, they would notice any time the player's gaze overlapped that point, while being less alert to other areas that the player was looking at. This meant that opponents who were deceived by an early decoy often remained distracted by it, even after the CA-GU player has moved on from the decoy. Below is a chronological example from Session 17 ($P17_B$):

> [03:09] *"She's really focusing on Pittsburgh."*
> [05:57] *"I'm still convinced she's going to go Pittsburgh somehow. She won't stop looking at Pittsburgh."*
> [08:39] *"She wants to do a loop down underneath. I just don't know how she plans to get to Pittsburgh after that."*
> [09:02] *"Ok now she wants to get to Pittsburgh from here" [after opponent claimed Helena to Denver route]*
> [11:49] *"I still think she is trying to get to Pittsburgh."*
> [12:46] *"Maybe she didn't need to go Pittsburgh at all." [after opponent started to take new ticket cards]*

Gaze misdirections and decoys became less effective as the board filled up with claimed routes, as these were taken as more reliable evidence of the players' intentions than the gaze tracking. To counter this, there are instances where CA-GU players even spent resources to claim a route along their decoy path, just to sell the idea that they were planning to build there.

Gaze Camouflage

Players used elements of the game board as cover for their real point of attention. This corresponds to *repackaging* in Bell and Whaley's typology [28]. This strategy involved the player

fixating for an extended period on routes that had already been claimed, to give the impression that they were considering this existing network, when in fact they were using their peripheral vision or brief side glances to plan a new path on nearby unclaimed routes. At other times this was done with elements outside the main game board, such as the cards around the outside. In some cases, the player appeared to be unaware that their face was turned towards the intended "cover point" but their eyes were still directed at their real goal, so the deception was unsuccessful.

Discussion

Players demonstrated a deep understanding of how to utilise gaze visualisation and their added knowledge about their opponent's plan. This went beyond simply discerning the opponent's intended paths and blocking them. A common tactic was to plan the timing of an action based on its proximity to the opponent's gaze: if the player saw the opponent looking in the vicinity of a path they wanted, that path became the player's first objective; other paths were de-prioritised relative to their distance from the opponent's area of focus. If the opponent's gaze was far away from all of the player's goals, the player was more likely to build up resources and bide their time before claiming any routes, allowing them to play in an efficient manner that avoided signalling their own plans to the opponent. When players were confident in their predictions, they were able to plan for deception. For example, $P9_A$ predicted a multi-step path that their opponent would take, waited until the opponent had built along most of that path (confirming the prediction), and then claimed the final section at the last moment, forcing the opponent to take a long detour after they had invested many resources into the original route. Gaze unaware players were much less likely to engage in such aggressive tactics, as they were less confident that their actions would be in the to stop or disrupt their opponent.

The benefits of the gaze visualisation were offset by a cost in players' time and attention. Players reported that the visualisation was distracting, and made it hard to concentrate on planning their own strategy: *"If my task is only to predict my opponent's moves, that would be substantially an easier task, especially with the eye tracking. But when you try to make predictions and play the game with a time constraint, the cognitive load starts to build up."* [$P2_A$]. Accordingly, we observed that players who could see the gaze visualisation spoke less about their own strategy and more about their opponent. Players who could not see the visualisation were more likely to

talk about the opponent's strategy in terms of potential risks and contingencies for their own strategy. Further, players speculated that the visualisation might be a net disadvantage in terms of winning the game due to its distracting nature, although most took a guess and mentioned that it would become more advantageous once players became accustomed to it.

The level of distraction may be partly attributable to the visualisation style. A few players commented that they would prefer a simple moving dot to the dynamic heatmap used in the study. The choice of the dynamic heatmap was informed by our prior study [188], in which it was the most accurately interpreted and second-most preferred visualisation style among nine formats compared. However, participants in that study were observers of a game rather than active players. It is possible that a less information-rich visualisation would be more manageable for players, who need to focus on developing their own plans as well as discerning those of an opponent. Moreover, the prior study showed substantial variations in preferences between participants. All in all, this suggests that it would be beneficial to give players some control over the gaze visualisation style, such as the ability to change its appearance or switch it on and off to manage the flow of information.

Non-live gaze visualisations could also help to avoid the distraction effects observed in this study. Mechanisms can be implemented to give each player a summary of where their opponent in an online match has looked, and for how long. For example, in *GeoGazemarks*, fixations were recorded and clustered to give a history of a user's points of interest to support free map exploration [90]. This is similar in function to the dynamic heatmap visualisation used in this study, although our heatmap provided a persistent summary only of gaze behaviour over the past few seconds; this persistent-summarising quality was one of the characteristics for which it was favoured in our prior study [188]. However, a non-live gaze visualisation would lose some of the affordances that provided for interesting gameplay opportunities found in our study, such as the use of *gaze scattering* strategy to distract an opponent.

While our study supports the general consensus of prior literature that gaze visualisations can be disruptive as well as informative, it also points to an opportunity for game designers to utilise this as a double-sided resource for players. We provide three examples of how this resource might be used as a "power up" in online multiplayer games. First, we can provide live information about where an opponent is looking, at the cost of distracting visual noise and the risk of being

misled by false signals. This can also take the form of non-live gaze visualisations, giving the player a summary of the opponent's gaze but leaving open the possibility that the opponent had manipulated their gaze to be deceptive. Second, we can limit where gaze visualisations appear; either activated by the player, activated arbitrarily (such as on a timer), or activated by other gameplay systems (for example, only appearing in areas that the player has previously explored, similar to *fog of war* in concept). Allowing the player to see gaze visualisations in only part of the game could leave them with a strategic choice of whether to seek further gaze information, at the potential cost of distraction, or allow themselves to be oblivious to their opponent's gaze. Third, we can use awareness of gaze sharing as a game resource, so that under certain conditions a player is notified when their opponent can see their gaze, allowing them to undertake deception strategies.

Generalising the implications of this study to other types of games should be done with caution, however. *Ticket to Ride* was chosen as an appropriate game for gaze sharing as it is turn-based and both players see a largely identical screen, so they can watch each other taking turns. These qualities are common to many digital board games, but are not present in many popular genres of online games. For example, in a real-time strategy game such as *Starcraft*, players scroll across a game map that is larger than the boundaries of their screen. A gaze visualisation for this kind of game would have to be anchored to the in-game location that a player is observing, rather than just the coordinates on the screen. This would reduce the amount of time each player spent observing their opponents' gaze, as it would allow them to monitor gaze activity only in areas they are concerned with, or only in areas they have previously explored (as in fog of war). A player may deduce, for example, that their opponent is not launching an attack as long as their gaze activity is isolated to the area of their own base. However, in a real-time game such as this, we assume the level of distraction caused by gaze visualisation would be even more acute than in a turn-based game as players do not have time to sit and think.

Players were successful in devising deception strategies that reduced their opponent's prediction accuracy about their plans, but also found degrees of difficulty in maintaining them (in line with [205]). All of the strategies we observed were conceived independently by the players, without suggestions from the researchers, and the stated inspiration behind many of the strategies was the player's own experience of observing the gaze visualisation. Players were largely successful

in converting their own experience into an accurate enough model of their opponent's thinking that they could implant a false belief about their strategy, or disrupt the prediction process in other ways. This demonstrates the application of *Theory of Mind*, as the *ability to deceive* is considered one of its hallmarks [151].

The effects of gaze visualisation were most pronounced in the early stage of the game. This was true both for the ability to make predictions and for the ability to deceive the opponent. At this stage, gaze data is a relatively clear signal, as there is little other information available in the form of routes claimed on the board. As the game progresses, deception through altered gaze behaviour becomes more difficult, as the accumulation of claimed routes on the board provides stronger evidence of intention than gaze behaviour. Claimed routes act as a costly signal due to the investment of resources involved, and so where there is a mismatch between routes and gaze (lack of correspondence, in Levine's terms [152]), the gaze information is ignored in favour of the route information.

This study extends prior work (e.g [188, 205]) by demonstrating that human subjects can make long-term predictions and employ deception given the ability to visualise gaze for longer durations. However, players commented that they lacked the attention and memory capacity to fully utilise the information that was theoretically available to them through the gaze visualisation. From our human observations, we found that players in our study ignored large areas of the board (e.g. whole left side) when they could see that their opponent was not looking at it. Based on their comments, it is apparent that this ability to triage the board on a wide scale was a substantial part of the competitive benefit players derived from gaze visualisation, comparable at least to the more fine-grained prediction of specific routes. We believe that future gaze-aware systems can do the same, by ignoring elements that are irrelevant to minimise computational cost. Our study also showed gaze enabled players to make earlier and more accurate predictions, which have positive implications for automated plan recognition systems [233]. Researchers have demonstrated that systems that can predict intentions based on gaze patterns [108], but beyond the question of how well computer systems can predict human intentions through gaze, there is the question of whether these systems could detect deception in human gaze behaviour, and distinguish "true" intentions from false.

Limitations

Whereas our previous studies have looked at gaze interpretation by passive observers (spectators) [188], this study involved gaze interpretation by active players. This allowed players to forestall their opponents' plans, potentially changing the outcome. As a result, a degree of human judgement was necessary to determine whether predictions were accurate at the time they were made, as players sometimes had to deviate from their original plans due to events in the game. Players used a combination of the gaze visualisation and in-game actions for their predictions, and we did not discriminate between guesses that were largely informed by one or the other.

During the study, players were prompted for their predictions only when they had been silent for a period of time, rather than at regular intervals; this helped to preserve the naturalistic flow of gameplay, but meant that prediction intervals varied between players and rounds. Asking for predictions at fixed intervals would standardise the results and allow a more standardised comparison between conditions and players, at the cost of a somewhat more artificial gameplay experience. Conversely, our results are subject to the experimental design—prompting players to make inferences about their opponents' strategy is likely to have influenced how they reasoned about the game, to at least a small extent.

Conclusion

In a lab study of an online strategy game enhanced with one-way shared gaze visualisation, we found that players were able use gaze visualisation to improve their ability to predict an opponent's intentions. Players were also able to alter their own gaze behaviour deliberately to hide their true intentions and convey false impressions, to a degree. We defined a typology of gaze-based deception strategies, which reflects the general theory of deception [28]. However, this deceptive behaviour required sustained effort, and players could not prevent some of their intentions from 'leaking' to the opponent through their gaze. Players found this gaze-based deception task to be challenging and engaging, and we have identified several avenues that may be fruitful for game design to explore.

End of Published Paper

5.3 Addendum: Further Analysis and Discussion on Deception

Due to the length limitations of the published paper, we excluded our comparative analysis of the gaze data collected under natural and deceptive conditions. In the *Experimental Setup* section of the paper, we stated that we logged the gaze data during the session for both players. Using this data, we can determine whether the gaze behaviours of players differed based on their awareness of being tracked by analysing its features. With reference to Figure 5.2, we compared the the gaze data from the *Gaze Unaware-Condition Unaware* (GU-CU) players in Condition 2 and the *Condition Aware-Gaze Unaware* (CA-GU) players in Condition 3. For this analysis, we have chosen not to compare using the gaze data from Condition (Baseline) from either player as they might still be learning how to play the game and, therefore, might have exhibited different eye movement behaviours such as longer fixations as they process new information visually.

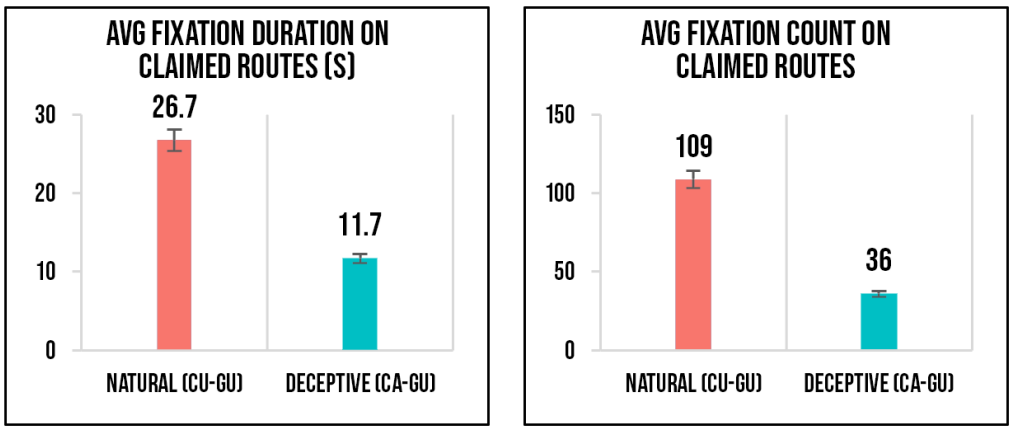


Figure 5.4: Deception Analysis. Comparison of fixation duration in seconds and fixations counts on ground truth for the natural gaze (CU-GU) condition and the deceptive gaze condition (CA-GU).

Our first measure compares the amount of time the players (CU-GU and CA-GU) on average looked at the routes they eventually claimed, while the second calculates the number of times they fixated on them. The rationale behind this measure is that in 'normal' conditions, players would look more at the routes they intend to claim, whereas players who were aware that their opponent is watching their gaze would avoid looking at the routes that are part of their plans. This comparison is to objectively show the types and magnitude of deceptive gaze behaviours, and to demonstrate whether deceptive behaviours can potentially be detected using features.

As shown in Figure 5.4 above, we found that on average, players not only looked less at their

routes when they were aware that their opponent was watching their gaze, but also less frequently. A Wilcoxon Signed-rank test shows that there is a significant effect of the awareness (*W=358, Z=3.547, p<0.01, r=0.56*) with players fixating on the ground truth routes longer when they were not aware that their opponent was watching their gaze. The median fixation duration for the CU-GU and GU-CA conditions were 27.2 and 11.2 seconds, respectively. We observed a similar effect for fixation counts. A Wilcoxon Signed-rank test shows that there is a significant effect of the awareness (*W=373, Z=3.88, p<0.01, r=0.61*). The median fixation count of CU-GU and GU-CA conditions were 89 and 32, respectively. These results support several gaze-based deception strategies where players avoided looking at routes and cities that were of interest to them—for example, *Gaze Averting* and *Gaze Scattering*.

Further, we analysed players' gaze behaviours on other *Areas of Interest*, for example, fixations on their own cards, the opponent's profile, their own profile, cards available for collection, non-ground truth routes, and gaze points that were outside the screen. However, we did not find any significant differences regarding fixations between the two conditions except fixations on the ground truth routes as highlighted previously. Although not statistically significant, we did note a marked increase in the length of time players spent looking outside the screen in deceptive condition than in the natural condition (30%, i.e. approximately 6 minutes versus 17%, i.e. approximately 3.5 minutes of average gameplay time). Such differences indicate strategies where players attempted to 'hide' their true intentions by not looking at the screen at all, such as in *Gaze Camouflage*. Then again, it is also possible that the players were looking outside the screen more as they were engaged in a conversation with the researcher.

The above results indicate that the monitored players exhibited altered gaze behaviours as did not want their opponent to know which routes they were considering in the deceptive condition. The acts of looking outside the screen more, averting their gaze (looking at their ground truth routes significantly less), and doing so significantly earlier than actually claiming the routes are at least three ways the players were deceptive using their gaze. These results also imply that players may have formed the intentions of claiming the routes reasonably well before acting on them, and this can also suggest that players in this game formed and tracked multiple intentions at the same time—for instance, the player planned for multiple routes for multiple tickets.

Figure 5.5 shows a comparison of low-level gaze features between the two conditions evaluated,

i.e., the average duration and counts for both fixations and saccades. The rationale behind this comparison is that players who are aware that their gaze is tracked would *glance* more on areas on the board, and therefore shorter fixations and more saccades. In our analysis, we did not find any significant differences between the two conditions, which was contrary to what we had expected. We believed that players would perform significantly more saccades in the deceptive condition, for example, when employing *Gaze Scattering*. This finding suggests that players looked around the board as "usual" but looked at other areas of the board more than the ground truth routes, as discussed above. A possible explanation is that *simulation* (overt, showing the false) strategies such as *Decoys* would produce similar low-level gaze features to actual intentions in the natural condition. However, this is difficult to analyse as players could have adopted multiple gaze-based deception strategies simultaneously, further, display natural gaze behaviour as when they play out their own strategic plans.

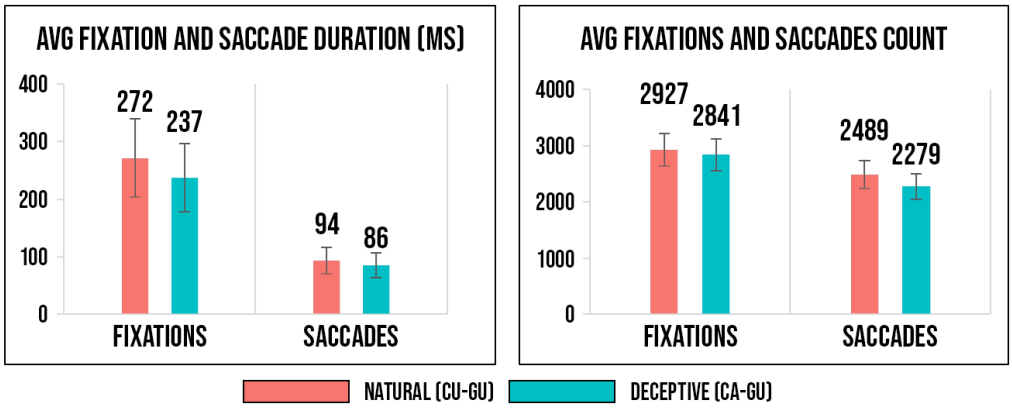


Figure 5.5: Comparison of gaze features between the natural and deceptive conditions.

Our findings so far show that we can potentially detect gaze-based deception, but this requires further investigation moving forward (e.g. comparing gaze patterns). Players in our study typically avoided looking at their true intentions (using both *dissimulation* and *simulation* deception strategies), and this is reflected in the data to some degree. On designing online systems that use human gaze for intention recognition, the likely assumption is that humans would behave rationally, including 'looking' at the interface or environment naturally. However, as shown, the awareness of being tracked can lead to changes in behaviour. Further, there are additional factors that could introduce uncertain and irrelevant gaze data, such as talking to other interlocutors or scanning the environment (no intentions), and this should be taken in to account as well.

5.4 Chapter Summary

This chapter demonstrates that gaze awareness can improve the capacity of human players to perform intention recognition in a competitive game setting—answering sub-research question *RQ1*. However, we found that although the human players in the study can improve their capacity to recognise the intentions of their opponent using gaze visualisation compared to when no gaze was shown, they are limited by their individual capabilities and can be subjected to gaze-based deception if their opponent was aware that their gaze was being watched. In majority of the games, the increase in either awareness led players to alter their strategies in ways that are detrimental to the player's own strategy. Moreover, considering that the gaze visualisation took their attention away from the game, we can argue that the increased capacity to infer intention using gaze visualisation may not be the best approach for revealing intentions. Nevertheless, gaze awareness added a novel aspect in our setting, and in general, players enjoyed the experience of being able to *see* what their opponents are looking to derive their intentions during gameplay.

Furthermore, employing gaze-based deception strategies also added an effortful and challenging aspect for players. The knowledge of being watched led to changes in gaze behaviour and strategy—*dissimulation* (hiding their true intentions) or *simulation* (conveying false impressions) or both. Then again, through our subjective and objective measures, we found that employing deceptive behaviours required sustained effort (in line with previous work [205]). Therefore, intentions were still 'leaked' to the opponent through their gaze, mostly occurring when players are planning or playing out their strategy. However, we note that it is often difficult to distinguish between true and deceptive behaviours, as 'deception detection' happens after the fact [152].

At the end of the previous chapter, we briefly mentioned that we have begun putting together a planning-based artificial agent that predicts the potential intentions of players based on the games used to generate the stimulus. After all, one of the thesis aims to increase an artificial agent's capacity to infer intentions by introducing gaze into its planning process for them to be better collaborators. In the next chapter, we will evaluate the robustness of the agent using both the data from the natural and deceptive conditions from this study as our next step. This transition is illustrated in Figure 5.1 in the introduction to this chapter, which summarises how this study informs the next part, Part 2, where we focus on the evaluation of the artificial agent.

The problem with the designs of most engineers is that they are too logical. We have to accept human behavior the way it is, not the way we would wish it to be.

– Donald A. Norman, *The Design of Everyday Things (1988)*

6

Study 3: Intention-Aware Gaze-Enabled Artificial Agent Evaluation

6.1 Introduction

This chapter presents the evaluation of the prototype of an artificial agent that combines gaze and observable actions to perform online human intention recognition. The prototype agent is conceptualised, built and evaluated on the data sets and knowledge accumulated from the two previous chapters which evaluated the performance of humans using gaze for intention recognition (Chapters 4 and 5). The development of the prototype agent is part of an ongoing collaboration between the *Interaction Design Lab* and the *Artificial Intelligence and Autonomy Lab* at The University of Melbourne, which aims at improving interactions between humans and agents through multimodal input[1]. Together, we initially published the model, along with its conceptualisation and initial results in the following peer-reviewed publication: [224], and later published an extended version the paper in the following peer-reviewed publication: [225].

In the previous chapter, Study 2, we showed that human players could increase their intention recognition capacity when shown the gaze of their opponent through the game. However, the additional layer of information increased their cognitive load, which resulted in no net gain in performance in terms of win rates. Hence, we propose that the task of gaze-based intention recognition could be delegated to an artificial agent instead, as we can argue that humans are limited in their capability to fully utilise the information that was theoretically available to them.

In the first half of this chapter, we present the implementation details of our prototype gaze-enabled artificial agent built as part of this thesis, including related work. In the second half, we evaluate whether the artificial agent can improve its capacity to recognise human intentions with gaze as with the human players, bringing us to our second sub-research question, *RQ2*.

> **RQ2:** Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?

As shown in Figure 6.1, this study employs the labelled data sets from both previous studies to answer RQ2. The data set from Study 1 served as a starting point for the construction of the agent model with promising results. The data set from Study 2 with multiple conditions allow us to evaluate the agent's capacity to perform intention recognition thoroughly, i.e., not only using natural gaze behaviours as priors but its robustness against deceptive gaze behaviours as well.

[1] https://cis.unimelb.edu.au/agentlab/human-agent-collaboration/

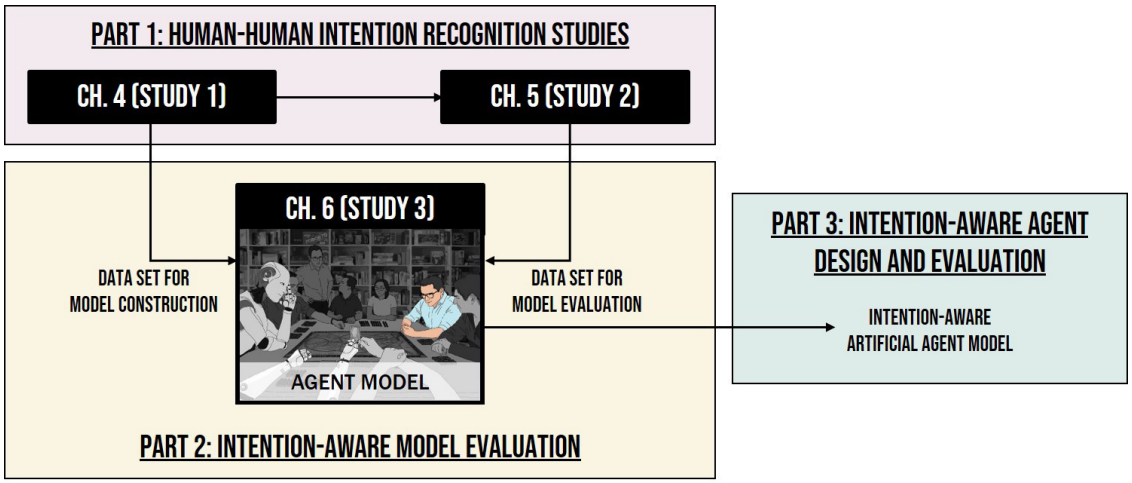


Figure 6.1: Part 2 (Study 3) focuses on evaluation of the agent model to be used in Part 3.

The two comparative evaluations performed in this study are as follows:

- First, we compared a model-based approach using only game-state data (*Model Only*) to the same approach enhanced with gaze-based priors (*Gaze+Model*). The results in Study 2 showed that gaze awareness could improve a human player's capacity to recognise intentions when compared to a baseline, where no gaze visualisation was shown. Here, we make the same comparison to evaluate whether this is the same case for the agent.
- Second, we compared the intention recognition performance of the agent against the performance results of the human players in Study 2 to determine if the agent outperforms the human player in recognising intentions. In Study 2, we found that players that are aware that their gaze is being watched would manipulate their gaze to hide their intentions, resulting in a lower intention recognition performance by the observing opponent. Hence, we further compared the performance using the data from both the natural and deceptive.

As with human players, the results show that intention recognition is possible from just inferring from the observable in-game actions along with the state of the game, but can be improved with gaze information (especially when no in-game actions have taken place). However, as discussed, humans players can be subjected to *confirmation bias* [191], especially they fixate on a prediction.

The first comparison showed that although the *Model Only* approach was reasonably successful at predicting intentions, the model with the addition of gaze as priors—*Gaze+Model*—resulted in not only more accurate predictions but also earlier predictions—answering *RQ2*. From our

observations in Study 2, we found that players ignored large areas of the board (e.g. whole left side) when they could see that their opponent was not looking at it. Based on their comments, it is apparent that this ability to triage the board on a wide scale was a substantial part of the competitive benefit players derived from gaze visualisation, comparable at least to the more fine-grained prediction of specific routes. Similarly, the artificial agent with access to the gaze data reduced its search space by'ignoring' the irrelevant elements and parts, which resulted in no additional computational execution cost despite the addition of gaze into its planning process.

The second comparison found that the agent outperforms the human players in performing intention recognition in both the natural and deceptive conditions from the previous study. We believe this is for two reasons. First, the agent has the sole task of providing the best plans possible using gaze as priors. Second, the agent can calculate features such as fixation counts and lengths, essentially filtering out the 'noise' from the natural constantly moving gaze point, which allows the accurate mapping on the areas of the game where the opponent is interested in. Hence, in the deceptive condition, it is likely that the true intentions of players still 'leaked' across, despite their attempts to manipulate their behaviours deceptively to hide them.

However, we note that the current iteration of the agent model, being an early prototype, does not perform well in all cases, such as when intentions are not sustained for long. Moreover, the evaluation of the agent's performance is based on the number of top plans it generates, in our case, the top 10 plans. This consideration is because the top plan might not be one that the human player is currently contemplating. During the game, the state of the game and the resources the players can obtain to carry our their plans influences their planning process, including what part of the plan they are executing and which plan. Therefore, it is common for players to kept their primary goal in their mind while constantly seek out alternate plans throughout the game. More importantly, this chapter demonstrates a working 'intention-aware' artificial agent with the capability to perform real-time gaze-based intention recognition for the next part, Part 3, where operationalise the agent for human-agent collaboration in a competitive gameplay setting.

In the following section, we further augmented our review of the literature on gaze-based intention recognition using machine learning approaches (Section 2.5) with related work that informs our model-based planning approach to enable our artificial agent to infer human intentions.

6.2 Related Work and Approach

Recent studies have successfully used automated planning for model-based intention recognition of intelligent agents [196, 204], including some preliminary studies that highlight its potential for human intention recognition in simple tasks such as shape drawing [245]. In these works, the recognition approach uses an automated planner to generate two plans for each potential intention: (1) a possible future trajectory that achieves the intention while corresponding to some sequence of already performed actions; and (2) the optimal plan for this intention that does not correspond to the already performed actions. Plans that are closer to the optimally-computed plans are deemed as more likely to be a rational 'fit' for the intention, and therefore more likely.

Earlier in Section 2.5, we covered several machine learning approaches for gaze-based intention recognition—all of which are viable approaches for our work. A model-based approach, however, presents an interesting and alternative way to perform intention recognition, that is, to use a model of the possible behaviours to perform a forward projection of candidate plans—starting from a sequence of recently-observed behaviours and to identify those plans that best fit the observations. Most existing works in this area use *plan libraries* for candidate plans, and assess how likely it is that the observed sequence of behaviours is a good fit for the prefix of such a plan. Many models of plan libraries have been investigated, such as Hidden Markov Models (HMM) [32] or belief networks [107] (see Sukthankar et al. [233] for a summary of existing approaches).

A downside of using *plan libraries* is that they require a set of potential plans for the other agent; which may be both prohibitively large, but also inflexible if the observed agent deviates from this set. Alternative approaches use concepts such as action models to temporally project possible agent plans into the future, essentially generating a set of candidate plans at runtime. Such models use belief-desire-intention plans [235] or planning models [204, 245]. For the purposes of this chapter, we refer to intention recognition as *planning* [150] interchangeably. Here, the set of candidate plans is not generated *a priori*, but instead, a planning tool is used to generate the set of possible plans that correspond to the observed prefix, and to assess their likelihood.

Ramirez and Geffner [204] solved the problem of needing plan libraries by generating two plans for each possible intention: the optimal plan for the intention in which the observed sequences of actions are satisfied by the plan; and the optimal plan for the intention in which the sequences are *not* satisfied. Their approach then determines the probability of seeing these observations

for the intention by measuring the *difference* between the two plans, and using Bayes rule to determine the probability of the goals, given some prior distribution over the goals.

More recent work simplifies this approach to reduce its computational burden. Vered, Kaminka, and Biham [245] extends Ramirez and Geffner [204]'s approach for *online* intention recognition. Instead of generating an optimal plan that does not satisfy the observations at each time step, they generate a single optimal plan for each intention, and then at each time step, generate a plan that satisfies the observations, and take the *ratio* between the cost of the original non-satisfying plan with the satisfying plans as a score, normalising these scores to get a final probability distribution. Masters and Sardina [166] independently make a similar observation for intention recognition in path planning, using the cost difference between the optimal plan that satisfies the observations and the optimal plan in general, thus also avoiding calculation of plans that do not satisfy observations. However, they show that for path planning (not task planning) can be calculated without considering the observations at all. They prove that the ranking of goals can be achieved by knowing only the agent's start point, potential goals, and its current location.

In this work, we hypothesise that incorporating gaze data to form a prior probability of these intentions can improve the agent's capacity to predict intentions, especially in terms of accuracy. Hence, our proposed approach is first to use automated planning techniques to generate candidate plans by observing the in-game actions of the observed human player. Then, as the human gaze is intrinsically linked to action [144], observing gaze over time can indicate their intentions (as found in the first part of this thesis). The agent can then combine the candidate plans and the in-game actions of the user are then combined with the user's gaze distribution (as priors) to generate a probability distribution of possible intentions.

There are several further reasons why we favoured a model-based planning approach for our prototype artificial agent. Firstly, aside from not having sufficient data to train machine learning models for use in our context, we wanted our approach to be generalisable to other scenarios and therefore chosen only low-level gaze features (e.g. fixation-based features) as a first step. Previous studies included high-level gaze features (or *interface-based* features e.g. total no. of visited areas) as part of their feature set, which requires the knowledge of the interface [26]. Hence, in theory, our proposed model only needs to know the coordinates of the objects in other contexts to work. Second, a model-based planning approach allows us to interpret how the

model made its predictions (i.e. 'white-box' approach), such as knowing which data that was used by the model to understand its decisions. Third, prior work has found several reasons why humans prefer working with an agent through planning, including the perceived reduction of cognitive load, and the ability to maintain situation awareness for short-term tasks [182].

6.3 Model

In this section, we outline our model-based approach, which consists of two independent components that form the input of our intention recognition algorithm: (1) the *gaze model* that processes the gaze information to determine the probabilities of different intentions; and (2) the *planning-based model*, which takes an action model and an observed sequence of actions, and determines the probability of intentions based on how well they fit the observed sequences.

6.3.1 Problem Formulation

We ground our problem on similar definitions based on intention recognition as *planning* [245]. Informally, there is a set of possible intentions that the observed agent can achieve, a set of actions that can be used to achieve it, and a set of observations we receive. The intention recognition problem is to determine the likelihood of the different intentions given the observations. In this context, observations are divided into two categories: (1) *ontic actions*; and (2) *gaze actions*. Ontic actions are those in which the agent under observation modifies the (physical or virtual) world [179]. Gaze observations are those in which the agent under observation is looking at 'regions' of the world, in which the regions are related to ontic actions. In our setting, gaze actions occur when a player is looking at different routes and destinations, while ontic actions occur when the player claims a route between two cities—which modifies the state of the game (or world).

Formally, an online goal recognition problem R is a tuple:

$$R = \langle W, s_0, I, A, G, O_a, O_g \rangle.$$

W is the world in which the agent operates, $s_0 \in W$ is the initial state of the observations, I is the possible set of goals/intentions, A is the set of ontic actions available to achieve the intentions,

G is the set of possible gaze actions, O_a is the sequence of ontic action observations (made up of actions from A), and O_g is the set of gaze observations (made up of observations from G).

This model aligns with Pacherie [195]'s model of intentions, who argues that there are three types of intentions:

1. **Distal intentions** (D-intentions), which are the long-term intentions of the agent and when achieved, generally terminates the agent's practical reasoning process,
2. **Proximal intentions** (P-intentions), which are short-term intentions that an agent derived to help achieve D-intentions,
3. **Motor intentions** (M-intentions), which are the grounded motor actions used to achieve D-intentions.

In this work, we have not considered M-intentions; however, we note that this model aligns with the notions of D-intentions and P-intentions: D-intentions are the set of potential *intentions* I, and P-intentions are the actions A. As such, the set of intentions in our work is the set $\Xi = I \cup A$. That is, an intention is either an intention to achieve an end state or an intention to execute some action in the future. This aligns with Bratman's notion that '*intentions that*' the world is in a particular state (a goal) expresses an intention around the world, while '*intentions to*' perform an action express an intention about an action [36].

A *solution* to the problem R is a probability distribution over the set of possible intentions Ξ indicating the likelihood of the agent fulfilling each intention, based on the observations O_a and O_g. For the remainder of the chapter, when we discuss intentions, we mean the set Ξ. From a modelling perspective, we can treat all intentions as actions, and have dummy terminal actions for each intention in I that correspond to that state being reached.

In this model, we assume a basic action model or STRIPS-like model of ontic actions [88], where each action is associated with a cost, have a unique name, have preconditions that determine under which worlds in W they can be executed, and effects that describe the changes made to W. Gaze actions have a *fixation count* that defines the number of times this gaze action was performed, and *fixation length*, which defines the time for which this gaze action was performed.

The computational problem is to determine how well the possible trajectories possible by sequences actions from A correspond to the observations seen so far (O_a and O_g). Assuming that

the actor aims to behave rationally, then intuitively, the similarity between the observations and the prefixes of the trajectories represents the likelihood of taking those trajectories—a rational agent will prefer cheaper plans over more expensive plans. From their P-intentions, we can infer the likelihood of their D-intentions. In this work, we further extend the related work to combine it with gaze data with the aim of improving the accuracy of the predicted intentions.

6.3.2 Intention Recognition

We divided our intention recognition approach into three steps, according to the different types of action observations. First, the approach uses the gaze observations to determine the probabilities of different intentions computed using fixation-based features. Second, independently of the first step, a model-based intention recognition approach is used to determine the probability of observing future trajectories of actions based on past observations. Finally, these two are combined to give final probabilities over the possible intentions. This separation of the two approaches supports online intention recognition that can incorporate new gaze data observations as they become available, without having to perform the expensive planning process again.

Combining Gaze and Planning

We first present the final step of our approach, which forms the backbone of the model. Our first observation is that the two types of observations can be treated independently, primarily because they describe two different types of actions. Gaze actions tell us about the possible future intentions that the human under observation is considering. Provided the person is not attempting to deceive the opponent using their visual behaviour; they will typically look at the regions that they are considering for future action more than those that they have already executed. In *Ticket to Ride*, the player would be more likely to gaze upon areas in which they are considering as possible routes, and when they perform the ontic action of connecting a route between two neighbouring cities, we can hypothesise that they are highly likely to look at that region immediately before connecting them. That is, they do not plan their actions blindly. Further, we assume a uniform prior O_g and note that techniques that use saliency of images can be used to obtain priors as well [115]. On the other hand, the ontic action observations are used to describe intentions *that have already been fulfilled.* Despite this, they can still be used

to predict future actions. In our setting, if the goal is to traverse entire paths (building a path connecting two cities), traversing a particular route indicates an intention that this will form part of a larger path, thus increasing the likelihood of the routes following the traversed route.

To combine gaze and ontic actions, we propose a simple model that treats the two types of observations independently initially and then combines them. However, we note the gaze and ontic actions are not entirely independent of each other: both are driven by the distal intentions (D-intentions). We model the gaze observations as priors over the chosen trajectories. The rationale behind this is clear: (1) a person is unlikely to act in any environment without sensing it first, thus gaze data related to particular intentions will almost always occur before any ontic actions; and (2) a person is less likely to look at parts of the environment with actions that they have already performed (unless that action can be performed again or is part of a bigger plan). As such, gaze actions provide a reasonable guide to future ontic actions.

Formally, we can describe the problem as trying to estimate $P(i \mid O_a, O_g)$ for each $i \in \Xi$; that is, the probability of an intention i given ontic and gaze action observations O_a and O_g. In this model, we assume that gaze actions from O_g are excluded once the related intention is achieved, thus implying that the probability of O_a is not influenced by the probability of O_g; although the inverse is not the case—past actions influences where people gaze. With this assumption, we can rewrite $P(i \mid O_a, O_g)$ as follows:

$$\begin{aligned} P(i \mid O_a, O_g) &= P(O_g, O_a \mid i) \cdot P(i) \quad \text{(Bayes rule)} \\ &= P(O_g \mid i) \cdot P(O_a \mid O_g, i) \cdot P(i) \\ &= P(O_g \mid i) \cdot P(O_a \mid i) \cdot P(i) \quad \text{(Assumption above)} \\ &= P(O_a \mid i) \cdot P(i \mid O_g) \quad \text{(Bayes rule)} \end{aligned} \tag{6.1}$$

Thus, the probability of a (P- or D-) intention i, given observed gaze and ontic actions O_g and O_a respectively, is the probability that the observed ontic actions would be taken if i was an intention, with the prior that i is an intention from the gaze data. Given this model, the remainder of the problem is how to determine $P(O_a \mid i)$ and $P(i \mid O_g)$.

For our gaze model, we employed two low-level (or basic) gaze features: *fixation length* and *fixation count*. Bednarik et al. [26] demonstrated that intention recognition could potentially

work well with a small subset of gaze features. Further, we note how 'gaze aware' players used the *real-time heatmap* visualisation in the previous study, whereby they observed of how long and how many times an opponent fixates on an object when predicting plans. Hence, the knowledge derived from such features could work for our agent as well, at least as a first step.

We define a simple model that uses these two measures to define the probability of each intention being a true intention. Recall from Section 6.3.1 that the gaze-based actions are summarised using fixation count and fixation length. A fixation is detected by the system when a human subject fixates on a target area for more than a threshold, for example, 60ms. A single fixation starts when the subject starts looking at a particular target (an item of interest) and then ends when the subject looks at another target. The change of the target areas is detected by tracking the *(x,y)* coordinates collected by the eye tracker. For each fixation, we maintain a fixation length, which is the difference between the start and end times of each fixation. For each target area, we maintain the total number of times a user visits the target area, and we refer to these counts as fixation counts. To cater for blinks (for periods shorter than 60ms), we merge the fixation lengths if the target being looked at remains the same after and before the blink.

The model of our data is assumed to be as follows. For each intention i, we have a pair: $\langle count_i, length_{i,j} \rangle$, in which $count_i$ represents how many times the gaze action corresponding to intention i was observed, and $length_{i,j}$ is a vector of j variables that represent the length of each gaze action. That is, $length_{i,j}$ represents the length of the j^{th} observation of i ($1 \leq j \leq count_i$).

We define fs_i, the *fixation score* for intention i as a weighted measure between fixation length and count:

$$fs_i = log(\lambda \cdot total_time_i + (1 - \lambda) \cdot count_i) \tag{6.2}$$

where $total_time_i$ is the sum of the fixation lengths for the intention i ($\sum_{i=1}^{i \leq count_i} length_{i,j}$), $\lambda \in [0, 1]$ is the relative weight given to the fixation lengths over fixation count. Because these are in different units (time and count respectively), it is likely that $total_time_i$ will be much higher than $count_i$. In practice, we have defined $count_i$ as the number of times i is looked at multiplied by the fixation threshold, so the two variables are on the same 'scale'.

Note the use of the log function in Equation 6.2. Using fixation length and count directly results in the linear function that increases monotonically. However, people's intentions change, and

they adopt new intentions over time. There are at least two options to mitigate this problem. First, discount the importance of earlier observations using a discount factor; or second, limit the increase in the importance of target areas as fixation length and count increase. By using the *log* function, we adopt the latter approach, which highlights important intentions while giving priority or importance to new intentions. We choose this approach because a person may stop looking at potential intentions once concrete plans have been formed, so discounting earlier intentions could miss important intentions.

Using the fixation weights, we define the probability of an intention i as its fixation weight normalised against other intentions:

$$P(i \mid O_g) = \frac{fs_i}{\sum_{j \in \Xi} fs_j} \tag{6.3}$$

where O_g is the set of fixation weights for the intentions representing the human gazing at targets that signal the intentions, and w_i is the fixation score for intention i.

Model-based Intention Recognition

Our model-based intention recognition is a simple generalisation of existing approaches of model-based intention recognition using planning. First, we define π_i as a plan that achieves intention i optimally while folding in observations O_a. That is, π_i is a sequence of actions from A, with prefix O_a, that executes P-intention i or achieves D-intention i, and where 'optimality' is defined by some criteria, such as length or cost of the trajectory. Note that there could be several such plans. Further, we define Π_i^{opt} as the *set* of optimal plans that execute/achieve i *without* (necessarily) folding in observations O_a.

We can now define the probability of observing O_a if intention i is a given trajectory that achieves i:

$$P(O_a \mid i) = \kappa \cdot \min_{\pi_i^o \in \Pi_i^{opt}} costdiff(\pi_i, \pi_i^o) \tag{6.4}$$

in which κ is a normalising constant and *costdiff* is a function that evaluates the 'cost' difference between two plans.

Informally, this equation states that the probability of seeing O_a if i is the intention is proportional to the cost difference between π_i and the closest optimal plans. Thus, a plan π_i that is closer to an optimal plan is more likely to have intention i. This definition may appear strange because O_a is a prefix of π_i by definition (π_i folds in observations O_a). However, we are not capturing the probability of O_a given just the plan π_i, but that the *intention* of π_i is to achieve i. Thus, when assessing π_i, if π_i is far from the optimal plan, then the person's true intention is probably not i.

We do not provide a complete definition of *costdiff* here; however, several such measures are possible, such as the simple difference in length, ratio of lengths, or the difference between the specific actions in the plans. There are a number of such functions proposed by other researchers (e.g. [204, 245, 166]). We use a domain-specific definition, based on ideas from Ramirez and Geffner and [204] and Vered and Kaminka [245].

Implementation

To implement such a model requires us to calculate $P(i \mid O_a, O_g)$ for every potential intention. This can be computationally expensive. Although our model combines the gaze and planning-based approaches into a single calculation (Equation 6.1), we note that a step-wise approach that first estimates $(i \mid O_g)$ from gaze actions can have practical value in cases where we may only care about the higher probability intentions.

Planning-based intention recognition is likely to be (relatively) more expensive than our gaze model because we must perform an expensive planning step for each intention, whereas the gaze calculation is performed as a simple weighted sum. To mitigate this problem, a simple step would be to use the planning-based model only on those intentions that have a high probability from the gaze-based model; for example, the top N intentions, or intentions with a probability above some threshold.

Equation 6.4 uses the set of all optimal plans Π_i^{opt} to evaluate how likely it is to see observations for a given intention. This step is clearly computationally expensive for any non-trivial systems. As such, approximations that, for example, compute only one optimal plan, may suffice in many settings. Alternatively, one could try to find the optimal plan that prioritises actions in π_i.

An updated version of the model presented above can be found in the following paper: [225].

6.4 Model Evaluation

6.4.1 Data Set

To evaluate the model, we utilised the labelled data set from the study in the previous chapter, where we recorded the gaze data and game states of participants pairs as they played the game of *Ticket to Ride* against each other under different conditions. For our evaluation, we specifically compare the data from the *Gaze Unaware-Condition Unaware* (GU-CU) players in Condition 2 (Gaze Viz.) and the *Condition Aware-Gaze Unaware* (CA-GU) players in Condition 3 (Deception).

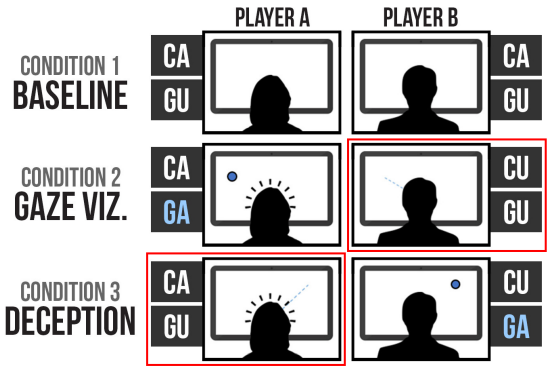


Figure 6.2: Labelled data set from Study 2. Each condition consists of the players' awareness of the study condition (**CA**: Condition Aware, **CU**: Condition Unaware) and their awareness of their opponent's gaze (**GA**: Gaze Aware, **GU**: Gaze Unaware). See Section 5.2 for more a complete description.

To elaborate, the CU-GU players in Condition 2 was entirely naive the condition, which gives us a data set of gaze behaviours under natural conditions with ground truths. We have previously evaluated the performance of the CA-GA players in Study 2, and therefore, we directly compare intention recognition performance between the agent and the CA-GA players in this study.

In Condition 3, the CU-GA players received the gaze visualisation of their paired CA-GU player but this time CA-GU player knew their gaze was being tracked. The CU-GA players, naive to this fact, performed gaze-based intention recognition without knowing that they could be intentionally misled. This gave us a data set of combined natural and deceptive gaze behaviours (players still look at things they were related to their goals). Therefore, as described in Chapter 5, though CA-GU players manipulated their gaze behaviour as they were aware that their gaze was being tracked, their true intentions still 'leaked' through.

6.4.2 Gaze Features and Planning

In this section, we briefly discuss some of the domain-specific details of our implementation.

Interface-Based Features: We divided the *Ticket to Ride* game window into labelled target areas according to areas of interest (see Figure 4.2), including the areas outside the shared map. The target areas signal the agent's intention. For example, if the player is looking at the cards, they may have formed the intention of picking up those cards; if the player is looking at a route, the player may have formed the intention of claiming that route. By considering the duration of the fixation and the number of times the player looks at a particular target, we make inferences about the likelihood of that intention being carried out. To establish which target was gazed at, we automated the mapping of gaze coordinates captured by the eye tracker to one of the target areas. The game state data for each game played was manually encoded this evaluation (e.g. cards drawn, routes claimed, scores and remaining resources).

Intention Model: We treated each route as a proximal intention (P-intention). Using this, we were able to construct a modified planning problem that consisted of only the routes that had been looked at, along with their corresponding probability. This represents the priors for the planning algorithm. Therefore, in this case, not only did we use the priors as probabilities, but also to reduce the size of the search problem.

Plan Generation: For each distal intention, the planner generated four different trajectories: two that fold in the observations and two that did not, corresponding to π_i and π_i^o respectively in Equation 6.4. We assume that there would be two strategies for fulfilling distal intentions: (1) build the shortest route; or (2) minimise the number of moves taken to build a route (that is, also including picking up cards, etc.). Thus, two plans that fold-in observations are generated — one for each possible way of achieving the distal intention — and two that do not — again, one for each possible way of achieving the distal intention.

For simplicity, we do not consider any other distal intentions that are possible in the game, such as claiming the longest route (which gains points in the game), trying to block other opponents routes, or actively trying to deceive other players on the intended route. The two former are the most common in our experience of playing and observing the games in Study 2.

To further simplify the process of planning, we assume that the naive player always has wild

cards. This means that we do not reason explicitly about the different cards the naive player has — something that is not observable and non-trivial to estimate. This not only reduces the complexity of the problem but also reduces the planning process to path planning.

Plan Cost Differences In Section 6.3, we did not provide a specific definition of the *costdiff* function. In our evaluation, we implemented a custom *costdiff* function. To generate candidate plans, we use Dijkstra's shortest path algorithm. However, players in the game can claim individual routes of a path in any order; that is, they do not need to claim the routes contiguously. Thus, we define the *costdiff* function based on the number of overlapping routes between the candidate plan and an optimal plan:

$$costdiff(\pi_i, \pi_i^o) = |\pi_i \cap \pi_i^o| \tag{6.5}$$

Note that π_i folds in observed ontic actions, which include the routes that are already claimed. The rationale for this that a path with a higher number of claimed routes (of an optimal route) should be scored higher.

6.4.3 Evaluation Setup and Performance Measures

The system requires a number of parameters. In Equation 6.2, $\lambda = 0.8$, that is we prefer fixation length more than fixation count when computing fixation scores. The standard fixation duration threshold was set to 60ms. This is the minimum duration a player has to be fixated on a single target area for the gaze activity to be counted as one fixation. The system generates a list of inferred intentions (routes), and for our evaluation, we only look at top 10 of the inferred routes, that is, the top 10 most likely routes the player is likely to choose.

In this study, we conducted two comparisons to evaluate the performance of the agent. First, we evaluated the difference in intention recognition performance with the addition of gaze data. The results in Study 2 showed that gaze awareness could improve a human player's capacity to recognise intentions when compared to a baseline. Here, we make the same comparison to evaluate whether this is the same case for the agent. We included a parameter to 'turn off' the gaze data in our implementation, which allows us to experiment with two different models: the *Model Only* system that makes inferences based solely on the model-based approach using

only game-state data and with no gaze data, which serves as a baseline for our study; and the *Gaze+Model* system, which is the same but incorporates the gaze data as priors into its planning process. This system builds a candidate list of plans based on these priors and the in-game player actions. We then compare the performance of each model using: (1) accuracy; (2) inference horizon; and (3) computation time. The measures are computed using the ten most highly-ranked inferences and also using only the most likely *plan* (sequence of actions).

Accuracy: We employed the same three measures as used in our previous studies to evaluate the overall success of inferring the ground truth, i.e., the actual moves made by players, rather than the intentions that they may have.

1. *Recall*: This measure is the ratio of the number of correct route inferences, and the total number of routes claimed (ground truth) in each game. When expressed as a percentage, this shows the success of the system in inferring the ground truths, i.e., the claimed routes.
2. *Precision:* This measure is the ratio of the number of correct route inferences and the total number routes inferred, which was fixed at 10 routes per inference.
3. F_1*-Score*: This uses the previous two measures and is computed using: $F_1 = 2 \times (precision \times recall) / (precision + recall)$.

For these three measures, an inference was considered correct if it appeared in the top ten most highly-ranked intentions, out of a possible 78. For analysis, we extracted the top 10 routes predicted by the models each time a route was claimed by the observed player and then calculated the accuracy at this stage. For each, we then computed the average accuracy, and finally, we computed and reported the average accuracy over the 20 games.

Inference Horizon: This measures how quickly a correct inference is made and is measured as the difference of time in seconds between when the system makes an inference, and a player makes the move corresponding to that inference and computing the averages of these times.

Computational Cost: This measures system execution time for inferring the top ten intentions.

Even though the accuracy measures do not consider where a claimed intention was ranked, we assert that this is still a useful measure, especially in the *Ticket to Ride* game, because players will have multiple possible intentions at any one point. Our aim is not to predict the next move, but the next N moves (in this case 10 out of a possible 78). In games such as *Ticket to Ride*, other

factors such as which cards the player holds impact whether claiming a particular route is even possible. As such, we assert that accuracy is a valid measure for this evaluation.

Second, we compared the intention recognition performance of the agent against the performance results of the human players in Study 2 to determine if the agent outperforms the human player in recognising intentions. Further, we found that CA-GU players in Study 2, who that are aware that their gaze is being watched, would manipulate their gaze to hide their intentions, resulting in a lower intention recognition performance by the observing opponent. Hence, we further compared the performance using the data from both the natural and deceptive conditions.

6.4.4 Technical Implementation

To evaluate our approach, we implemented the artificial agent using Python plus several Python libraries such as *networkx*[2] to compute the shortest path between two points. We then utilise the *matplotlib*[3] and *shapely*[4] libraries to visualise the planning process of the agent in real-time (e.g. Figure 6.5) and for displaying features (e.g. fixation count and length on cities and routes).

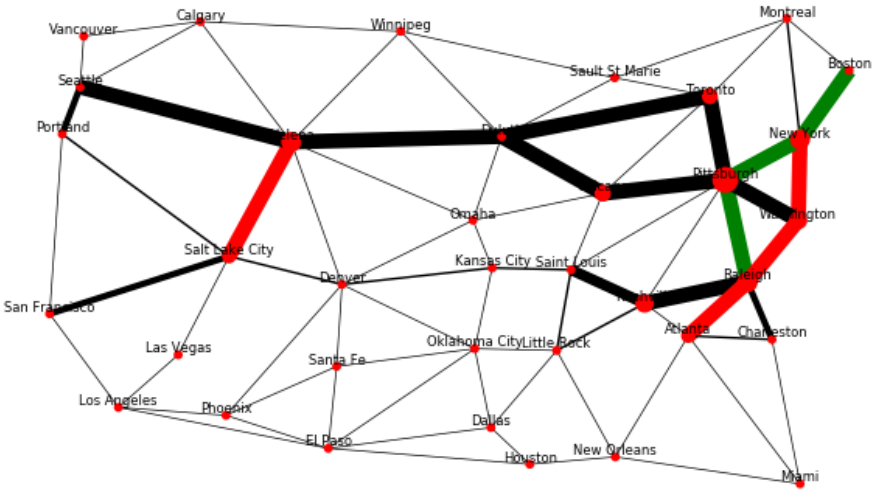


Figure 6.3: Example of opponent plans predicted by the agent. The black lines represent possible intentions of the opponent while coloured lines represent claimed routes (green: player, red: opponent). Thickness of the black lines represents the likelihood of the opponent claiming a particular route.

Figure 6.3 above shows a visualisation of the planning process of the agent, where we have replicated the map of *Ticket to Ride* using a graph representation. In this example, we show

[2]https://pypi.org/project/networkx/
[3]https://pypi.org/project/matplotlib/
[4]https://pypi.org/project/Shapely/

the top 10 most likely plans of an opponent during one of the games predicted by the agent. Each plan, *i*, is generated using Equation 6.1. The thickness of each route reflects the number of times it appears in the top 10 plans. If we consider top 10 plans, we are most likely looking at a common set of routes and some variations of a plan that a player is considering in the game. Here, the opponent is going towards the west coast, and either going to Seattle or San Francisco but via a common set of routes going towards west. When building systems that will convey the intentions or plans graphically, it makes sense to consider not only the most likely plan but also top D plans that can capture both, the common actions as well as contrasting ones. While we do not evaluate the effectiveness of this approach, it appears to be reasonable, at least in our setting.

6.5 Results

6.5.1 Performance Comparison Between *Model Only* with *Gaze+Model*

In this section, we discuss the performance of *Gaze+Model* system when compared with the *Model Only* system. We tested the effects of the recognition approach on the dependent variables with a Welch's t-test. We tested the data for normality using Shapiro-Wilks tests and did not find any significant violation of normality.

For this comparison, we evaluate the overall success of inferring the ground truth, and we considered all inferences made up to the point in the game when the player completed their first ticket, as opposed to the end of the round or once they completed all tickets. We did this for several reasons. First, all players completed at least one ticket, but not all players completed all their tickets in the data set. For players that completed all tickets, they may not have display strategic intentions after due to the absence of goals. As observed in previous studies, their strategies may most likely switch to tactics to reduce their opponent's score. Second, players on average completed their first ticket 12.4 minutes into the game ($SD = 4.9$, $min = 3.1$, $max = 19.5$), about partway through the game round (mid-game), and therefore sufficient data for us to determine if the model can predict (early) intentions. The route built to complete the first ticket often integrate part of the routes required to complete the other tickets, and therefore the remaining tickets get completed in a shortly after the first. Third, to simplify the evaluation process due to the complexities from the game itself. As players usually have multiple intentions

(or plans) to complete their goals, players may switch goal completion priorities, such as they intended to complete Ticket A, but an opportunity arises to complete Ticket B first, or that the players might have abandoned certain plans entirely due to factors outside their control.

Accuracy

We analysed whether gaze data had a positive impact on inference accuracy. Table 6.1 shows the *Precision*, *Recall* and F_1*-Score* for the two systems when using top ten, five and the most likely route respectively. With a Welch's t-test, we found a significant effect of the intention recognition approach (*t(19) = 5.09, p < 0.01, Cohen's d=1.14*) with the *Gaze+Model* approach offering a higher accuracy (71%) than the *Model-Only* approach (47%).

	Precision		***Recall***		F_1-***Score***	
N	*Gaze+Model*	*Model Only*	*Gaze+Model*	*Model Only*	*Gaze+Model*	*Model Only*
N=10	0.71	0.47	0.50	0.34	0.56	0.38
N=5	0.63	0.35	0.44	0.24	0.49	0.28
N=1	0.55	0.23	0.37	0.16	0.42	0.18

Table 6.1: Proximal intention prediction for the two approaches.

Figure 6.4(a) shows the percentage of successful inferences of the ground truths when considering the top 10 potential routes, sorted by the *Gaze+Model* accuracy scores for presentation purposes. Clearly, gaze data had a positive impact on the accuracy of the inferences, scoring higher on accuracy in 16 of the 20 games, and scoring equally on three of the remaining four. Game 2 is the only game in which the *Model Only* approach outperformed the *Gaze+Model* approach, while in games 1, 8, and 13, the accuracy are the same for both approaches.

While these results could be just statistical anomalies, we note that in these games, it appears that the players change their intentions rapidly and employ several different strategies, compared to the other games. Some collect many cards before claiming routes while others claim routes as soon as they have required cards. Further, in game 2, it is clear that the players are trying to block the intended paths of each other, which forces a change in intention (and lack of a sustained intention). Path blocking is an intention that was not encoded in our planner. It appears that the *Model Only* system can tackle such situations slightly better than a system that incorporates gaze data, because in the *Gaze+Model* system, the gaze data may be noisy and thus not useful.

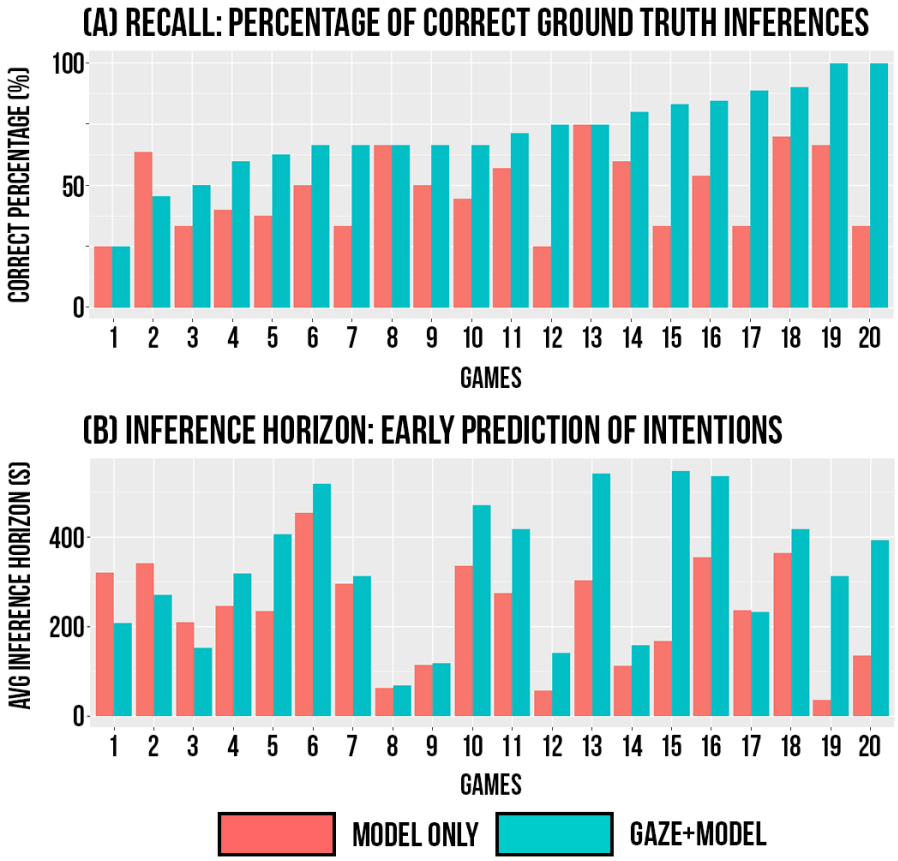


Figure 6.4: Comparison of accuracy and inference horizon results for the two models.

Inference Horizon

In this measure, we tested whether the agent with gaze data enabled the agent to perform early (or fast) intention recognition. Figure 6.4(b) shows the inference horizon for the two approaches, with a higher inference horizon implying an earlier prediction. These results show that overall the *Gaze+Model* system was able to predict the intentions quicker than the *Model Only* system, doing so in 17 out of the 20 games. Overall the average prediction horizon of the *Gaze+Model* system was 327 seconds (approx. 5.5 minutes) while the horizon for the *Model Only* system was significantly lower at 233s (approx. 3.7 minutes). This early prediction is illustrated in Figure 6.5.

With a Welch's t-test, we found a significant effect of the intention recognition approach ($t(19) = 3.31$, $p < 0.01$, *Cohen's d=0.74*) with the *Gaze+Model* approach recognising intentions earlier (327s) than the *Model-Only* approach (233s). For reasons outlined in Section 6.5.1, in games 1, 2, and 3 the *Model-Only* approach performs better. The post-game analysis reveals that blocking results

Figure 6.5: Visualisation of artificial agent performing early prediction using with gaze.

in rapidly changing intentions and players tend to act almost immediately. As this chapter's purpose is to determine whether the agent can improve its intention recognition performance with gaze, we leave the exploration of strategies to mitigate these instances for future work.

Figure 6.5 exemplifies a game where the agent performs well early in the game and, also at different intervals. In this example, the agent has already made some predictions solely based on gaze data, just under a minute into the game. A minute after, at two minutes, we can see that the agent became more confident in the prediction once the player (opponent) has claimed a single route in the south (in red). At about ten minutes into the game, we can see that the agent did in fact successfully predict the overall plans of the player very early in the game through their gaze.

Computational Cost

This measures shows that the cost of the two approaches was similar. With a Welch's t-test, we found no significant effect of the intention recognition approach (*t(620) = 0.83, p = 0.41, Cohen's d=0.14*) with the planning times of the *Gaze+Model* approach ($M = 76ms, SD = 40ms$) being similar to the *Model Only* approach ($M = 79ms, SD = 24ms$), a positive result in our case.

Overall, the results so far showed that incorporating gaze into intention recognition was effective, at least in *Ticket to Ride*. The results show that adding gaze-based priors to model-based intention recognition more accurately determined intentions, determined those intentions earlier, and at no additional cost; all compared to a model-based-only approach. However, the results also show that improvements can be made, particular, when the players' intentions changed rapidly.

6.5.2 Performance Comparison Between *Human Players* and *Artificial Agent*

In this section, we compared the intention recognition performance results from the data set against the performance of the agent to determine if the agent outperforms the human player in recognising intentions. In Study 2, we found that players that are aware that their opponent is watching their gaze would manipulate their gaze to hide their intentions—resulting in a lower intention recognition performance by the player. Hence, we further compared the performance of the agent for the deceptive condition as well.

	Human Players			**Artificial Agent (Top 10 Plans)**		
Condition	*Precision*	*Recall*	F_1*-Score*	*Precision*	*Recall*	F_1*-Score*
Natural	0.48	0.34	**0.39**	0.74	0.59	**0.64**
Deceptive	0.41	0.28	**0.31**	0.60	0.59	**0.59**

Table 6.2: Performance comparison between human players and artificial agent for the natural and deceptive conditions.

As shown in Table 6.2, the artificial agent is not only able to make more accurate predictions than the human players in the natural condition but also in the deceptive condition. We ran a Welch's t-test for both conditions, and we found a significant effect of between the human players and the artificial agent for the natural condition ($t(19) = 4.98$, $p < 0.01$, *Cohen's d=1.11*) and the deceptive condition ($t(19) = 5.03$, $p < 0.01$, *Cohen's d=1.13*). The artificial agent performed better in 16/20 games in the natural condition and 18/20 games in the deceptive condition.

These results suggest our approach of using a model that combines gaze and ontic actions is robust to deceptive, at least in the context of the *Ticket to Ride* game. The results also suggest that gaze is intentional, and in our study, players still signalled their game-related intentions even though they attempted to deceive their opponents through various *gaze-based deceptive strategies*. To some extent, this could be related to the time- and resource-related pressures or opponents' moves that may have to lead the players to let their 'shields down', and thus to give away their true intentions as they still needed to carry out their own plans.

Though we did not empirically investigate the contribution of the ontic actions to the success of the *Gaze+Model* model, the results indicate that the success of the model can be attributed to the use the non-deceptive ontic actions. This means that players must eventually signal their intentions with ontic actions if they want to achieve their goals, and being too deceptive with their actions would cause the player to waste valuable moves and resources.

6.6 Chapter Summary and Conclusion

In this chapter, we present a novel approach that combines gaze and model-based online intention recognition to infer the intentions of human subjects. Our approach uses gaze data to build probability distributions over a set of possible intentions, which are then used as priors in a model-based intention recognition algorithm (P-intentions to make inferences of the D-intentions). The performance results show that gaze-based priors significantly improves the accuracy and quickness (horizon) of the inferences when compared with classical model-based approaches.

Moreover, when designing systems that may use human gaze to infer intentions, most often one assumption is that people would behave rationally using their gaze. Some factors could introduce uncertain and irrelevant gaze data, such as talking to other interlocutors, scanning the environment, or being actively deceptive. Hence, we also evaluated our approach using under deceptive conditions as well. The results show that the artificial agent performs better than human subjects in both natural and deceptive conditions, especially when combined with gaze. Further, gaze allows the agent to exclude some of the goals entirely from the search space resulting in no additional computational cost, similar to the human players in Study 2 that ignored parts of the board where the opponent did not gaze upon. However, this process must be considered carefully as the experience of people may influence their gaze behaviours: human do not always look at target areas before acting on them, and they may not necessarily look at the keys on the keyboard when typing, for example. In future work, we can also improve the agent handles the dynamic nature of intentions, as they can change rapidly. We can also improve the model to either try to filter out older, noisy intentions or mitigate their effect, would, and extend the range of possible interaction modalities for gaining information, such as gesturing.

One of the main advantages of our approach is that it is essentially 'white-box', that is, when the model makes a prediction, we can look up or track up at the underlying data used by the model to understand the agent's predictions. This makes not only communicating predicted intentions possible but also explaining the reasons behind these predictions possible. For example, in our setting with the *Ticket to Ride* game, when the agent predicts the opponent will claim a given combination of routes from one city to another, a human collaborator can query about the reasons behind predictions, especially if communicated by the agent, such as, how much or how often the opponent has gazed upon the on routes, and what has been claimed so far.

Over time, AI will evolve to include more collaboration-based applications, with multiple human-machine teams working together.

– David Martinez

7

Study 4: Assistive Intention-Aware Artificial Agent Language Identification

7.1 Introduction

The findings of this chapter have been published as part of the following peer-reviewed paper:

Joshua Newn, Ronal Singh, Fraser Allison, Prashan Madumal, Eduardo Velloso and Frank Vetere. 2019. Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents. In *Human-Computer Interaction – INTERACT 2019.* (INTERACT '19). Springer International Publishing, Cham, 255-281. DOI: https://doi.org/10.1007/978-3-030-29384-0_17

The paper in its original published format is found in Appendix A.3. The paper includes two studies, and this chapter presents the first of the two.

In this chapter, we focus on the communication protocols of the artificial agent, which will allow the agent to communicate its inferences to an assisted-player as a first step towards the human-agent collaboration. In the previous study (Chapter 6), we evaluated the performance of a prototype artificial agent that incorporates gaze input into its planning process for online human intention recognition. The evaluation shows that the agent can not only increase its intention recognition capacity when we incorporate gaze into its planning process but that it also outperforms the human subjects in our earlier study for both natural and deceptive conditions (in terms of accuracy and timeliness). While the artificial agent can infer the intentions of a human subject, it lacks communicative capabilities required for the predicted intentions to be useful to a human collaborator. By employing the agent as a collaborator in our setting, the agent can assist the human player by performing the cognitively demanding task of intention recognition on behalf of the team. We hypothesise that this *division of labour* would allow the player to better focus on forming the best strategy possible given the information from the state of the game and the intentions of other players from the agent. The agent is therefore required to *interface* and *interact* with the assisted-player through effective communication protocols.

Hence, in this study, we identified a preliminary language for communicating intentions that our enhanced artificial agent could adopt to assist a human player in our setting. Early intentions derived from gaze has shown to be beneficial for strategy formulation and given that the agent can perform intention recognition earlier and more accurately than a human collaborator, we can start to design complementary capabilities for supporting human-agent collaboration (RQ3).

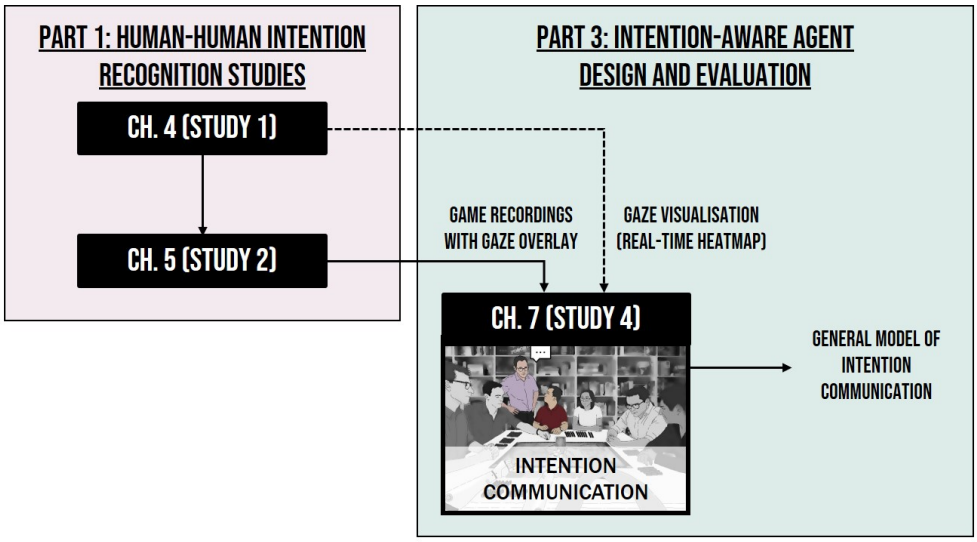


Figure 7.1: The *real-time heatmap* visualisation from Study 1 and the game recordings in Study 2 are used to elicit how humans communicate about intentions in Part 3.

Figure 7.1 above illustrates the progression of our aims in giving our prototype artificial agent the ability to increase its intention recognition capacity with gaze so far. For the human-human intention recognition studies performed in Part 1, we employed gaze visualisations as a way to utilise gaze input in our setting. Study 1 found that live gaze, represented as a *real-time heatmap*, works well for increasing a human observer's capacity to perform intention recognition.

In Study 2, employed the gaze visualisation throughout the game and observed that the human players could not only keep track of their opponent's intentions through their gaze but also able to explain and reason about them. In this study, we take on a similar approach as with Study 1, where we employed the use of human spectators, and instead of short clips, we use the full game recordings of the *aware players* from Study 2. Whereas Study 1 focuses on their ability to recognise intentions, this study focuses on eliciting how humans can proactively communicate the intentions of others to a teammate to improve their awareness of the situation, given the ability to perform gaze-based intention recognition, forming RQ3.1 below.

> **RQ3.1:** How should an artificial agent communicate intentions given the ability to perform gaze-based intention recognition?

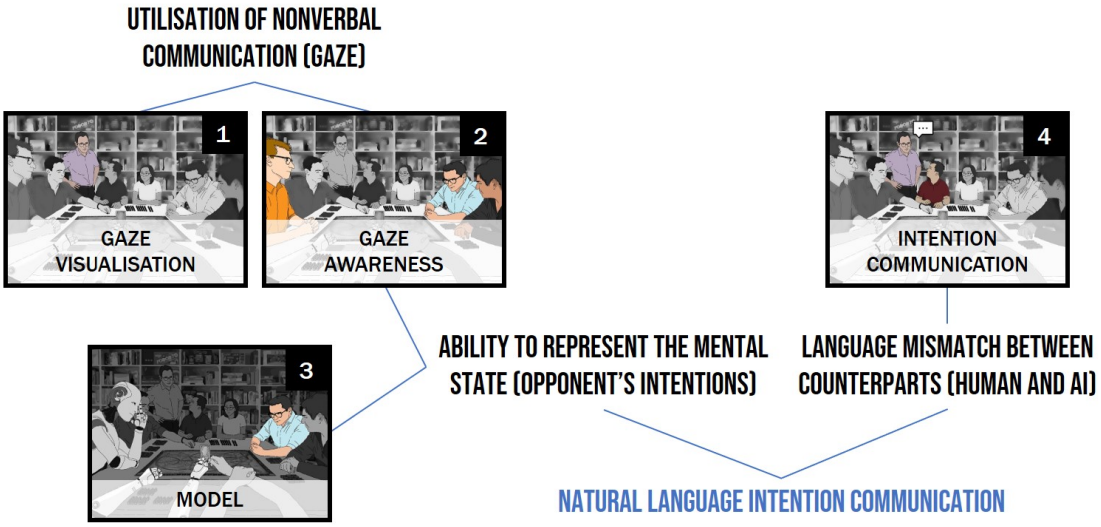


Figure 7.2: Relationship between the *human-centred prerequisites*, studies and outcomes.

Figure 7.2 shows the outcomes of the previous studies in line with Licklider's *human-centred prerequisites* outlined in Chapter 2. In this study, we address the *matching the language between the counterparts* prerequisite, where we provide the agent with the ability to communicate with its human collaborator through natural language about the intentions of their opponent. So far, we have addressed two other prerequisites for our setting. In the first two studies, Study 1 and 2, provided us with insights into how our prototype artificial agent can utilise gaze input. In Study 2, the players with access to gaze information were able to keep track of their opponent's intentions, to some degree, and therefore able to represent the mental model of a tracked player through their gaze. In Study 3, we showed that the prototype agent could also represent the mental model of a tracked player when we incorporate gaze into its planning process.

This study, Study 4, explores how agents should communicate information about a human player given its ability to perform gaze-based intention recognition. In essence, whether the agent can abstract the key information from its mental model of their opponent's plans, and communicate them to the assisted-player in an effective manner. From the study, we determined the linguistic aspects of communicating its observations based on implicit gaze information and explicit in-game actions, and subsequently, how predictions can be explained in a manner that is useful for a human collaborator. This study further contributes to the rapidly growing field of Explainable AI (XAI), where we show that observable nonverbal behaviours can be adopted as a basis for explaining human behaviours and intentions through a human-centred approach.

7.2 Related Work and Approach

To build the communicative abilities of the agent, we must first consider how an ideal agent would communicate intentions once recognised while taking into consideration the optimum level of support as different levels of artificial agent support can result in changes in cognitive workload, positively or negatively [43]. Hence, this study also focuses on the *what* and *when* aspects of intention communication, in addition to the *how*. At this stage, we have chosen to a support a human collaborator by communicating the intentions of an opposing player as a first step, and not on providing recommendations as this requires the agent to know the plans (or intentions) of the human collaborator. As providing the agent the ability to provide recommendations based on intentions requires significant exploration on combining different mental models (and how best to communicate the recommendations of the player to the agent), it is currently beyond the scope of this thesis and will address this as part of future work (see Section 9.4).

Another important consideration is the artificial agent's capability to explain its reasoning process to foster trust in the context of human-agent collaboration [110, 216]. In real-world human-human teaming, a human collaborator can produce some sort explanation on how they arrive at a decision (or inference) when asked, which makes explanations a crucial aspect of communication and decision making in collaborative settings. This capability to explain is evident in the first two studies; where participants could verbally explain their reasoning process about their inferences or about a move they made just made when prompted. Therefore, we extend the intention communication model to include explanations to augment the agent-derived predictions for two primary reasons. First, users would often seek an explanation of its output when interacting with an agent to improve their own decision making [247]. Second, agents that provide explanations supports transparency and in turn, may help foster trust in the system [100, 161, 174, 192, 249]. In our setting, it is especially vital to include explanations as the agent can produce *early* predictions, that may not immediately be obvious to the assisted player.

Gaze awareness has also been shown to be especially beneficial for situation awareness in social settings. Situation awareness is defined by Endsley [78] 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*". In our setting, players use various sorts of information to make a strategic decision, which includes not only their observations of the state of the game

past and present, but also future projections. For example, *"If my opponent is going to do this, then I will do this." or "I think my opponent might do this, so I will do this instead."*. In the previous study, we show that our prototype agent can perform plan projections better with gaze input in terms of accuracy and timeliness. Therefore, we hypothesise that the more accurate and timely awareness obtained from collaborating with the agent would allow a human player to improve their ability to assess the situation correctly while simultaneously narrow down the multiple predicted intended plans a player typically maintains during gameplay. Hence, the design of the artificial agent in this thesis is aimed at providing the best possible predictions at any given time to reduce communication overhead so that the player can then focus on strategy formulation.

At the same time, the prototype agent supports transparency as we have chosen to adopt a *white-box* approach. This approach allows us to interrogate the data or features to form an explanation for each prediction communicated. Further, this approach allows us to adopt the three levels found in Situation Awareness-Based Transparency Model (or SAT model) [58] to support the user. In this model, the agent communicates different types of information at each level to support a human collaborator's situation awareness. 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. Following the SAT model, we can design the communication for our agent using two levels, the middle level for enabling explanations (reasoning) and the top level for communicating predictions (projection).

Furthermore, for our agent to provide effective explanations, it must be at the right level of detail [131]. 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, hence decreasing its effectiveness [192]. Aside from the *presentation format* (how to explain) and the *explanation content* (what to explain) [75, 131] of an explanation, it is also important to consider the application domain and the audience of the explanation [98]. Hence, it is essential for us to simulate our study in our chosen competitive gameplay setting.

To examine RQ3.1, this study uses a variation of the Wizard of Oz technique to elicit how a collaborator who can infer intentions using the combination of gaze and in-game actions would communicate the intentions to a teammate. In a typical Wizard of Oz study, a researcher secretly

plays the role of the computer system while a participant interacts with it [132, 207]. In our study, we indirectly asked participants 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 manner. The use of human elicitation was inspired by previous gaze-enabled human-agent interaction studies (e.g. [202]), and a similar approach has been used in the context of machine learning to 'bootstrap' a dialogue system on human-generated activity [208]. The Wizard of Oz technique has also been used to revealed important insights regarding the usage of natural gaze behaviour for object manipulation tasks in multi-display environments [21].

7.3 Study

7.3.1 Participants

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.

7.3.2 Experimental Setup and Procedure

Upon arrival, we sat the participant in front of a computer setup and obtained their written consent to participate in the study. As illustrated in Figure 7.3, the participant and experimenter sat on opposite ends of the table so that the experimenter's display was not visible to the participant. The setup consisted 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 are two other players in the next room preparing to play a game *Ticket to Ride* against each other, and that they had been 'randomly selected' to play the role of a 'predictor-explainer' (or appraiser), whose role is to watch the game between the two other players via a live stream and to send assistive messages to one of them in secret, their "teammate".

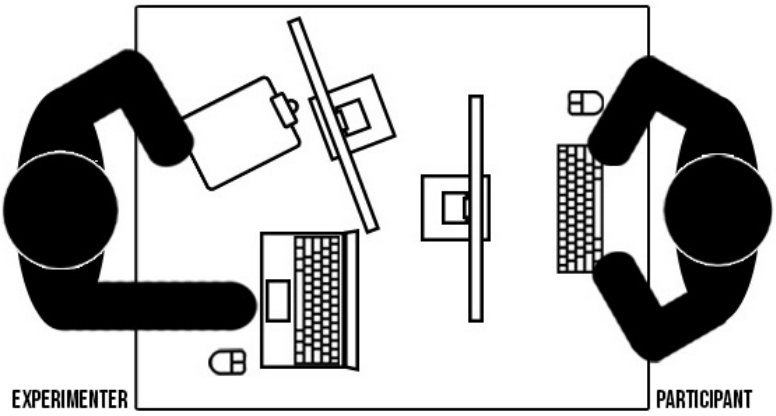


Figure 7.3: Study 4 'Wizard of Oz' Experimental Setup.

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 Study 2 recordings. The player whose gaze was doing was naive to the fact their gaze was being visualised to the other player, meaning they natural gaze behaviours were displayed (GU-CU condition). We favoured 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 of this study.

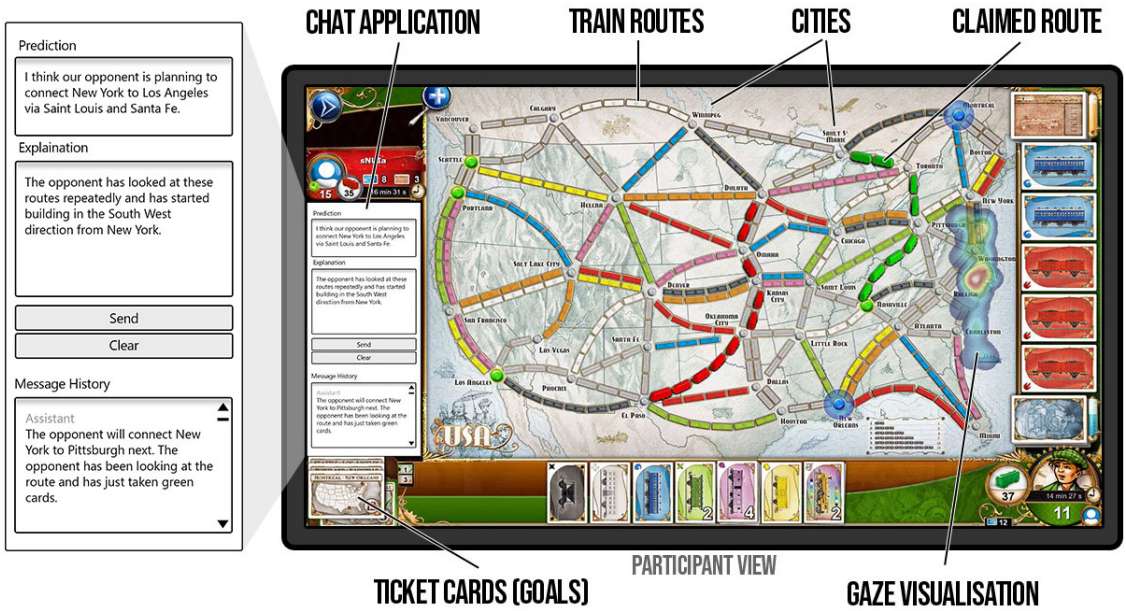


Figure 7.4: Study 4 Participant View.

Figure 7.4 shows the layout of what participants saw on their screen. We simulated a live stream using the game recording and told them that the 'live' dynamic heatmap visualisation is the gaze of the 'opponent' player. For this study, we designed and employed a protocol to continually reinforce the participants' belief that they were engaged in a live online game with unseen two opposing players in the next room throughout the study. For example, as we designed for each session to last a maximum of an hour, we informed the participant in advance that the game would begin at a fixed time, such as partway through the session, as all "three" participants (including themselves) 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.

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). We selected clips four varied intention recognition scenarios clips of *Ticket to Ride* with the live dynamic gaze visualisation overlay (taken from our recordings in Study 1). The purpose of this preview was to enable the participant to start thinking about how they could form *predictions* from the information available from the game and the gaze visualisation, and then from *explanations* in text form about their reasoning process. This step also 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 unseen "teammate" would find helpful for their strategy and to build the teammate's trust by being transparent in how they derived their predictions by providing explanations.

Next, we demonstrated a simple chat application that served as the means of communication with their unseen "teammate" (see Figure 7.4). 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", "Hmmm ok").

At the prescribed start time (usually about 30 minutes into the study), the experimenter began streaming the recorded game as if it was a live game feed and informed the participant that the game had started. During the study, 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.

7.4 Study Findings

We elicited a total of 249 raw messages ($mean$ = 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 recommendation(s) 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 predictions. Finally, we obtained a total of 246 messages after our filtering process for data analysis.

7.4.1 Prediction Format

For the prediction part of each message, we stripped them into its essential and 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 11 key prediction formats (see Table 7.1), 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 [17, 94, 156, 232], we will introduce uncertainty when communicating predictions in our next study, including stating alternate routes when the likelihood of the plan is similar (e.g. To [City] or [City] through [City] from [City] or [City].).

Pattern	N
from [City/Area] to [City/Area]	130
from [City/Area] to [City/Area] through n[City]	37
to/towards [City/Area]	35
interest in/along/around [Area]	15
[Intentional Action]	8
around [City]	5
to [City/Area] through [City/Area]	4
between [City] and [City]	4
from [City/Area]	3
from/to [City] to/from [City] or [City]	3
from [City] to [City] in/to [Area]	2

Table 7.1: Key Prediction Formats

7.4.2 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 the findings of a study that employed 'expert explainers' (shoutcasters) [69]. 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 for us to build a general model, we turn to Malle and Knobe [164]'s explanation model for labelling the properties for more complex explanations elicited from our participants 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*, in which we defined as O_A, and *Intentional Action*, in which we defined as I_A. As our chat 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" [117], becomes a valuable source of information for participants throughout the game, especially when only a few observable in-game actions have been performed by the opponent.

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

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]."

7.4.3 Reasoning and Communication Strategies

Participants adopted two general strategies for reasoning about and communicating the predicted plans of their common 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 [124]—*System 1* (heuristic, intuitive) and *System 2* (systematic, analytical)—also called 'thinking fast' and 'thinking slow' respectively. 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 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.

As expected, participants on average generated more predictions at the beginning of the game and fewer predictions towards the end of the game—reflective of plans being formed. 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 that it is often difficult to tell how far in advance the opponent will carry out the prediction (e.g. in how many turns).

7.5 Chapter Discussion and Conclusion

This chapter explores how artificial agents should communicate with a human collaborator given the ability to improve their capacity to infer intentions through gaze observation. We employed an 'inverted' Wizard of Oz technique to elicit how humans *should* communicate intentions of an opponent and subsequently explain their reasoning process given the ability to infer intentions with gaze visualisation. In Study 2, we observed that humans could explain their reasoning process and as the evaluation of the prototype agent shows promising results, we shift our focus onto building the agent's capacity to communicate for use in a human-agent collaboration setting. This enables the agent to be collaborative, as it can assist a human player in decision making by providing situation awareness of the opponent's future states on behalf of the team.

In this study, we determined the linguistic aspects of communicating inferences based on implicit gaze information and explicit in-game actions, and subsequently, how predictions can be explained in an assistive manner, resulting in a preliminary language model of intention communication. This model forms the basis for a computational model of explanation, where we can map nonverbal behaviours onto explanations, and therefore generalisable to some degree to other settings. The properties and strategies for intention communication are reflective of the findings in previous work (e.g. spatial, temporal and quantitative [69]). The study finds that participants in their 'predictor-explainer' role also portray uncertainty when unsure about their predictions, an important consideration for natural communication. This study further contributes to the rapidly growing field of Explainable AI (XAI), where we show that observable nonverbal behaviours can be adopted as a basis for explaining human intentions and behaviours.

In the next chapter, we evaluate the usefulness of the communication protocols found in this study as part of a broader study—where we evaluate the use of the agent during live gameplay.

Talent wins games, but teamwork and intelligence
wins championships.

– Michael Jordan

8

Study 5: Assistive Intention-Aware Artificial Agent Real-Time Interaction Evaluation

8.1 Introduction

The findings of this chapter have been published as part of the following peer-reviewed paper:

Joshua Newn, Ronal Singh, Fraser Allison, Prashan Madumal, Eduardo Velloso and Frank Vetere. 2019. Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents. In *Human-Computer Interaction – INTERACT 2019.* (INTERACT '19). Springer International Publishing, Cham, 255-281. DOI: https://doi.org/10.1007/978-3-030-29384-0_17

The paper in its original published format is found in Appendix A.3. The paper includes two studies, and this chapter presents the latter of the two. In this chapter, 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 [109, 224], 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.

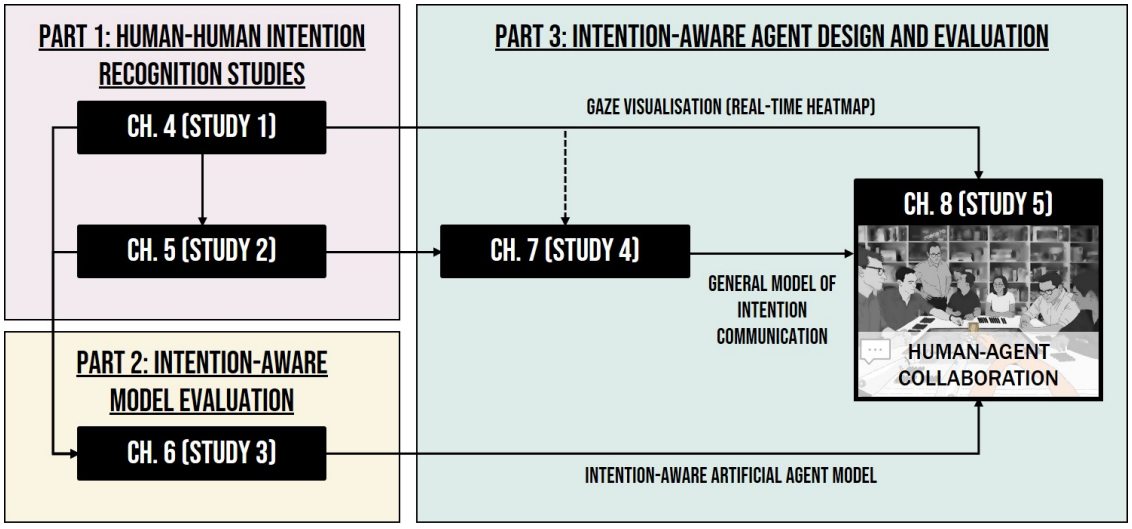


Figure 8.1: The figure illustrates how the outcomes of the previous studies converge in Study 5, where we operationalise the artificial agent in our chosen human-agent collaboration scenario.

As shown in Figure 8.1, the outcomes of previous studies converge in this study—Study 5—the final study conducted in the thesis. From Study 2, we recognised that humans are limited in their capability to not only infer but also to keep track of intentions, especially with the addition of gaze due to its cognitively demanding nature. Gaze provides a continuous stream of information about a player's interest in our scenario. However, this continuous stream that makes it difficult to differentiate between its behaviours, which subsequently contributes to additional cognitive processing beyond keeping track of what has been looked at. For these reasons, we proposed that it might be better suited to delegate the task of gaze-based intention recognition to an artificial agent, which is one way we can effectively divide the labour between the team. Hence, it has shaped the subsequent studies, particularly with Study 4 that focused on the agent's ability to communicate the intentions of others to a human collaborator. Therefore, with all the necessary components in place, we can begin to evaluate the use of the artificial agent for collaboration in our setting in real-time to improve decision making, which brings us to RQ3.2 below.

> **RQ3.2:** Does delegating the task of intention recognition to the agent improve the player's decision-making?

We examine this question through Study 5, which focuses on the novel collaborative capacities of the intention-aware artificial agent, whereby the agent can provide situation awareness with the capacity to explain them by performing intention recognition on behalf of the human-agent team. In turn, the agent can potentially reduce the cognitive load of the assisted human player and therefore allow the player to form better strategies along with an increased awareness of the situation. This study further evaluates the experience of players using an artificial agent that can project future states for effective strategy formulation afforded by its ability to create awareness through abstracted intention information and its ability to provide accurate and timely predictions on intentions. Hence, the study compared the experience of users of using the gaze visualisation found in our first study as our baseline condition against the use of the agent, where we provide predictions proactively at two levels, abstract (prediction only) and detailed (prediction with explanation) informed by the previous study—Study 4. By comparing these conditions, we can further determine the elements that are missing when we begin to abstract intentions from the rich gaze information using gaze visualisation.

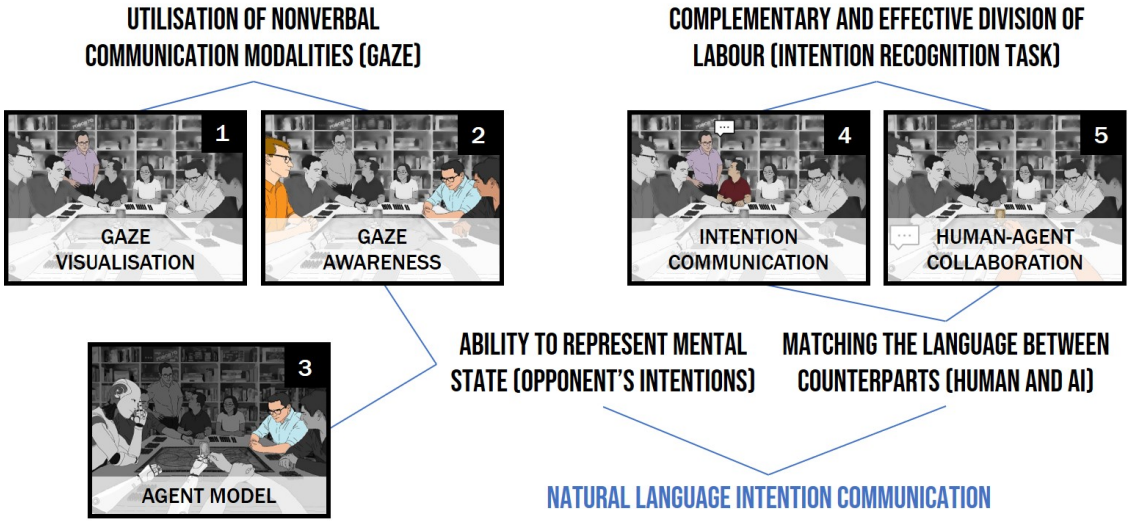


Figure 8.2: The outcomes of each study in relation to the *human-centred prerequisites*.

Figure 8.2 above shows how the four previous studies have addressed, to some degree, the different human-centred prerequisites outlined in Chapter 1. As shown, Study 1 and 2 helps enforces the importance of using nonverbal communication (i.e. gaze), with respect to intention recognition in a competitive gameplay setting. Study 2 and 3 demonstrates that gaze affords the ability to represent the mental states of a human player, in our case, the intentions of an opponent. In Study 2, players kept track of the intentions by annotating on a map of the game, while in Study 3, we produced a similar but improved graph where we can observe the likelihood of predictions through the agent. In both cases, we can visually represent the real-time plans of the opponent and how it changes over time (as illustrated in Figure 6.3).

In Study 4, we address the language mismatch between the human and the agent, by investigating how humans should communicate intentions using natural language, given the ability to predict them using gaze. Study 5 combines and simultaneous evaluates the two models arising from Study 3 and 4 to form components of an 'ideal' intention-aware agent—the *artificial agent model* and the *language model*, respectively. We designed both models from the ground up to work in tandem to improve the agent's capacity to collaborate with a human player in our setting. The study presents the first step towards using artificial agents that have the ability not only to interpret intentions but also communicate to a teammate. We demonstrate that by introducing gaze input for intention recognition, we can improve the capability of artificial agents to assist in improving the human-agent team's overall situation awareness, and therefore decision making.

8.2 Study

By combining the *language model* derived from our previous chapter with an instance of an intention-aware artificial agent that we evaluated in Chapter 6, we can now evaluate the experience of playing an online strategy with and without agent assistance.

In this study, we also evaluate the model while comparing the existing approach of using gaze visualisation to infer intentions to our proposed approach of abstracting the intentions from gaze input into written predictions through an artificial agent. We, therefore, designed a within-subjects user study with three conditions, in which we provide varying levels of information to an assisted player. In the first condition, we provide players with a live 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 is 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. In the second and third condition, 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. This interaction approach is akin to 'AI-advised human decision making', where the system provides information and the human makes the final decision based on the information past and present [22].

8.2.1 Intention-Aware Gaze-Enabled Artificial Agent

We instantiated an artificial agent that performs intention recognition using the combination of ontic actions and gaze using the planning-based model, detailed in Chapter 6. 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 [182]. 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 the previous chapter was for three reasons. First, using the data set from Chapter 5, the agent scored positively higher compared to a human interpreter using gaze visualisation ($F_1 - Score$: 0.64 versus 0.39 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 goals of the opponent (destination cities) to the wizard, the system remains far better at discriminating and predicting the most likely plans and can provide this information earlier as well. This capability ensures relative consistency of predictions and provides a realistic impression of what such systems can do across the players. Third, as found in Chapter 5, human interpreters can be subject to biases, especially when the human interpreter fixates on incorrect predictions and overlooks other predictions.

8.2.2 Experimental Setup

Figure 8.3 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 attached to the bottom of their screens; the evaluated player (P_A) was equipped with Tobii Pro X2-30[1] (for pupillary data), and 'naive' opponent (P_B) was equipped with a Tobii 4C eye tracker. The bottom part of the figure shows the rule-based assembly process. To assist in the assembly, we use the graph visualisation that displays the predictions made by the agent developed in Chapter 6. In summary, 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, i.e. alternate plans.

[1] https://www.tobiipro.com/product-listing/tobii-pro-x2-30/

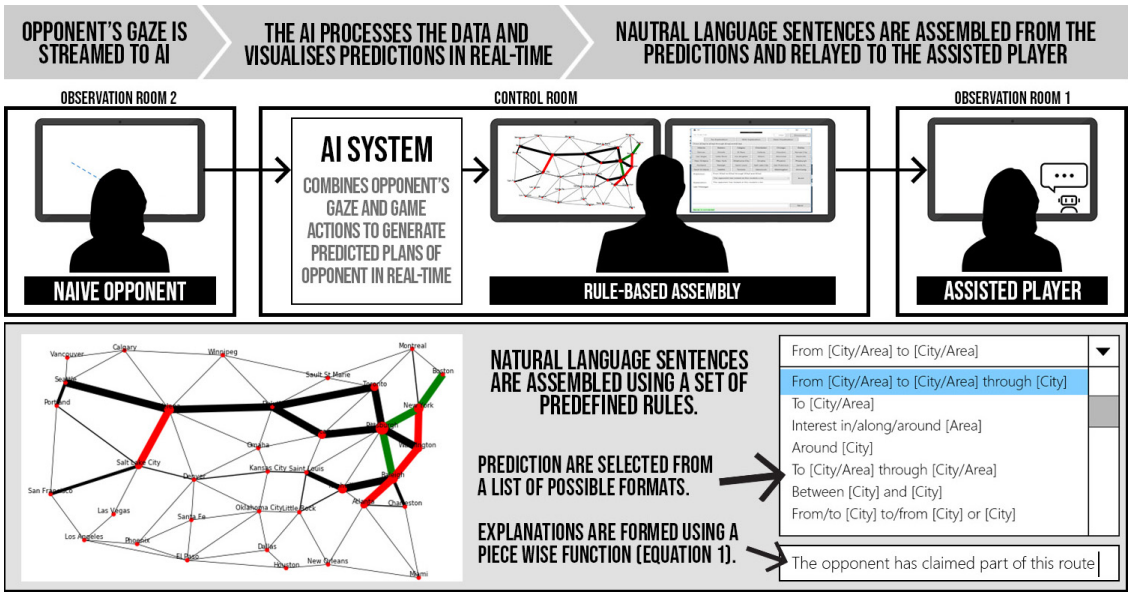


Figure 8.3: Top: Experimental Setup and Communication Flow. **Bottom**: AI System Visualisation and Assembly Process. Opponent's intentions are displayed by increasing the line thickness of routes. The thicker the line, the more likely the route will be claimed. Coloured lines represent the claimed routes (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).

8.2.3 Study Conditions

We designed three conditions to evaluate three levels of information abstraction (see Figure 8.4). At the lowest level of abstraction (GAZE VIZ), we show the assisted player (P_A) the gaze of the naive opponent (P_B) using the *real-time heatmap* gaze visualisation. This first condition allowed players to make their inferences on their opponent's plans at the cost of their attention and serves as a baseline condition as we displayed the visualisation throughout the game.

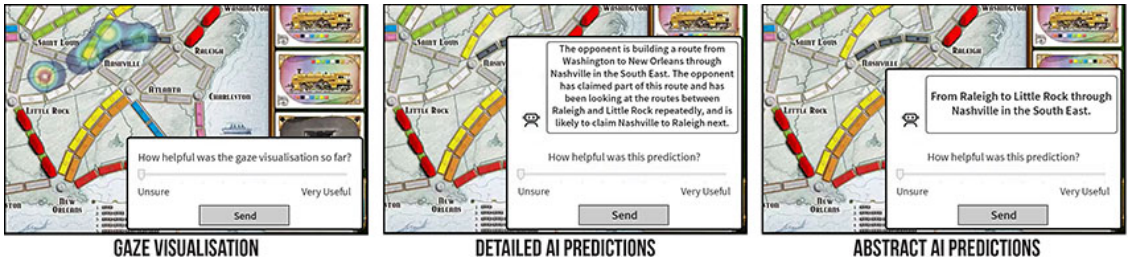


Figure 8.4: Study Conditions.

At the mid-level (DETAILED AI PREDICTIONS), we assemble the intentions and observed behaviours using the preliminary language model formed in the previous study. Here, we presented the prediction as an *Intentional Action* [164]—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 PREDICTIONS), the agent provided an abstract predicted plan through one of the predictions formats in the language model. Figure 8.4 shows an example of how all three conditions are represented based on the same underlying data.

As we based the AI prediction levels on the reasoning and communication strategies described in Section 5.2, we simulate the agent's communication frequency accordingly. For the DETAILED AI PREDICTIONS condition, we require the formation of detailed messages and therefore set the frequency to every 2 minutes so that the agent system can make a sufficient observation to form the best possible prediction and explanation. For the ABSTRACT AI PREDICTIONS condition, the frequency was set to a minute (60 seconds), as we only need to send the best possible prediction at that point in the game. We counterbalanced the three study conditions using a Latin square to minimise any learning effects. As this is a within-subject study, we only subjected the conditions to the assisted player (P_A), who is the 'aware' player between the two. The researcher facilitating promotes the conditions as 'modes of assistance' at the start of the session. For both AI-assisted conditions, the researcher made explicit that the artificial agent uses both the opponent's gaze behaviour and observable in-game actions to generate the predictions on behalf of the team.

8.2.4 Measures and Analysis

We designed a repeated-measures questionnaire to evaluate the experience of the assisted player in and across each condition (found in Appendix B.2). As there was no existing questionnaire specifically designed to measure the experience of 'intention awareness', we formed our questions based on our questionnaire employed in Chapter 5 (found in Appendix B.1), which was used to measure 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 included questions to measure the assisted player'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 two conditions with agent assistance, we included two additional measures, which

questioned 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 PREDICTIONS condition, we asked players about the *clarity* of the explanations (agent's reasoning process) to validate messages formed using the language model. At the end of the study, we measured the overall experience of using all three condition and 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 post-study semi-structured interview.

To measure cognitive workload unobtrusively, we used the recently proposed Index of Pupillary Activity (IPA) metric [72], 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. [43]), such as *long fixations* (fixations that are longer than 500 ms), which indicate deeper cognitive processing. We also employed the popular NASA-TLX questionnaire [99] to capture perceived workload based on six subscales—mental demand, physical demand, temporal demand, performance, effort and frustration.

8.2.5 Participants

We recruited 60 players (34M/26F) for the study and allocated them randomly into two equal groups according to gender (17M/13F in each) from The University of Melbourne. 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 for post-study analysis. 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. The naive players were necessary to simulate realistic gameplay with natural eye movement behaviour during the study. All players were compensated with a $20 (AUD) gift card for their participation.

8.2.6 Procedures

To manage the complexities of the study, all three researchers involved in the experiment followed a strict rehearsed protocol. Both players were given an initial briefing together in the same room

upon arrival. We explained that their physiological signals will be tracked throughout the study for post-study analysis. We then provide players with the plain language statement, consent form and basic demographic questionnaire to fill out before separating randomly into one of the two observation rooms at the University's usability lab with an assigned facilitator. Each player was then instructed to play the game's interactive tutorial for up to 10 minutes to get familiar with the game and its controls, regardless of experience. Both 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 game 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 assigned rooms, the facilitator (F_A) informed the assisted player (P_A) 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 gaze using the eye tracker's default calibration before starting the tutorial. We then informed the player that they would play three rounds of the game against their opponent (i.e. the naive player - P_B) and will receive 'additional information' about their opponent's intentions without their knowledge, which will vary according to the condition.

For all three conditions, the player received prompts with a slider below each prediction (see Figure 8.4). For the GAZE VIZ condition, the prompt appeared every 60 seconds for consistency. The primary purpose of the slider is for players to reflect on the additional information presented to them and to give an impression that their feedback is taken into account. We also instructed the player to verbalise why they had given a particular rating. At the end of each condition, we administered the NASA-TLX questionnaire immediately before asking them to fill up the

repeated-measures questionnaire on their experience about the round they just played. This ordering was intentional as their subjective workload may change after filling up the repeated-measures questionnaire. Once completed, the facilitator 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 player's gaze using the eye tracker's before playing the tutorial. At the end of each condition, we administered the NASA-TLX questionnaire and a Games Experience Questionnaire (GEQ - Core Questions) [39]. The primary purpose of both the questionnaires was for the player to fill up the time while player P_A went through a longer post-study questionnaire and interview. Any gaps in time were filled up by facilitator F_B, who will engage in a conversation about the game round they just played.

8.3 Results

The first part of this section presents the overall results from our various subjective and objective measures, as previously outlined in Section 8.2.4 (Measures and Analysis). In the second part, we present and discuss the experience of the players with and without the agent from the insights provided by the post-study semi-structured interviews in relation to our various measures. Figure 8.5 summarises the median scores for the responses from the repeated-measures questionnaire.

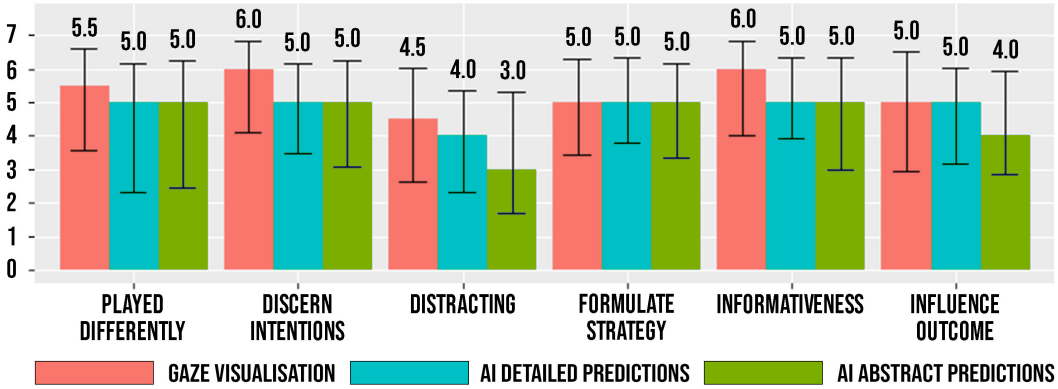


Figure 8.5: Study 5 Questionnaire Results (Repeated-Measures Questionnaire).

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 conditions with agent assistance 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.

	Preferred			**Demanding**			**Useful**		
Condition	*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 8.1: Post-Study Ratings for Each Condition.

Table 8.1 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 the post-study semi-structured interview as players were asked to reflect on their reasoning behind their given ratings. Table 8.2 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 PREDICTIONS conditions ($W=0$, $Z=4.78$, $p<0.05$, $r=0.87$), and between both the DETAILED AI PREDICTIONS and ABSTRACT AI PREDICTIONS conditions ($W=0$, $Z=4.78$, $p<0.05$, $r=0.87$).

Condition	*IPA (Hz)*	*Avg. Long Fixations (>500ms)*	*Avg. Sacc. 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 8.2: Results of Cognitive Workload Measures.

The results show that assisted players on average had longer fixations in the DETAILED AI PREDICTIONS condition, which could simply be because they needed time to process the predictions. We found no significant differences between GAZE VIZ and ABSTRACT AI PREDICTIONS conditions, which suggests that players did not require a longer time to parse the predictions in the ABSTRACT AI PREDICTIONS condition, indicating that the ABSTRACT AI PREDICTIONS case did not introduce any significant workload while achieving similar awareness as the GAZE VIZ condition.

Though the agent did not necessarily decrease the cognitive workload as compared with the current approach of gaze awareness through visualisation, the NASA-TLX questionnaire scores indicated that players perceived the 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 VIZ condition was higher when compared to the AI conditions, a Kruskal-Wallis test showed no significant differences among the three conditions.

8.3.1 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, similar to the gaze visualisation in Study 2. For example, 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. From [City] to [City] in the 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 visual component.

In our analysis, we found that assisted 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 sent each time, it took them further away from their current task. Between the AI conditions, players preferred the abstract AI predictions over the detailed AI predictions in general as the messages were more concise and therefore needed less time investment in deciphering them and subsequently utilising the information:

> **P8**$_A$: *"I liked the simplicity of the information it [the artificial agent] gave me, it was very easy to filter."*
> **P10**$_A$: *"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."*
> **P23**$_A$: *"Shorter and brief hints were easy to understand and helpful."*

There were overarching reports that predictions from the agent 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.

8.3.2 Gaze Visualisation

A third of players (10/30) explicitly mentioned that the gaze visualisation was 'distracting' (such as 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 them to be less focused overall.

About 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—similar to our findings in Study 2. 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 players of 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 the repeated-measures questionnaire responses, as shown in Figure 8.5.

Table 8.1 shows that although gaze was found to be most demanding, it was rated most preferred and useful. There are two possible explanations for this response. 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. This could mean that players just needed to know roughly where their opponent's intentions are, as opposed to the specifics. 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 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 from the gaze visualisation. In comparison with the AI conditions, the information was only available when the agent communicates, leaving the players on occasion with no or minimal additional information. From our observations throughout the studies, players usually make use of the time to analyse the information and to plan accordingly. When the opponent took a long turn, players in the AI conditions had to wait for the agent to send them predictions.

Without the ability to initiate the interaction with the agent, the player has to make use of the information that was then available to them, such as from the past agent's predictions (that may have changed) or past in-game actions of the opponent.

8.4 Discussion

In this chapter, we evaluated the prospects of using an ideal intention-aware artificial agent in a human-agent collaboration setting. In our case, an artificial agent that assists a human teammate with the inferring intentions (or plans) of an opponent in a competitive gameplay setting. Through a within-subjects study, we evaluated the responses of 30 players that played *Ticket to Ride* against a naive opponent with and without agent assistance. As the task of gaze-based intention recognition was found to be cognitively demanding in our previous studies, the study compares the use of the agent against using gaze visualisation for intention recognition.

The overall results showed that the agent could facilitate awareness of another user's intentions without adding visual distraction to the interface; however, the cognitive workload was similar across all three conditions presented, suggesting that the manner in which the agent communicates its predictions requires further exploration. This means that the agent neither impacted the players' performance nor decreased the cognitive workload as initially hypothesised. The game itself might have introduced cognitive workload, but it 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, most likely due to the reduced distraction (periodical abstracted information vs. continuous information), but also because there is no continually moving gaze overlay over the game. Therefore, we acknowledge that irrespective of the mode of communication, any task requiring the processing of information generates cognitive workload.

On the other hand, the subjective assessments indicate that the agent was successful performing gaze-based intention recognition and communicating them to the assisted player in a way that matched the informativeness of the gaze visualisation. In turn, the predictions and explanations provided by the agent early in the game allowed the assisted player to formulate better strategies.

Further, the 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 note that an agent-assisted approach can potentially scale well for multiple users in which the agent can determine what is

the most relevant information to communicate, whereas 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 and explanations predictions. Hence, while textual representations of intention served as a good starting point for our work, it was insufficient to improve the player's performance (in terms of decision making). We look towards the use of *visual explanation* as our future direction, where agents can use multimodal (e.g. pointing, voice) or multimedia approaches (e.g. annotations, animations) to communicate intentions effectively.

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 knows about the opponent, 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 [202], 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 Chapter 6, and shows that nonverbal inputs such as gaze can be used as a basis for generating 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. As such, 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. [241]'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 [146, 159] involves both a cognitive process to derive an explanation and a social process of communicating the explanation to a human [159, 174], there is a clear scope of expanding our approach to generalise our findings to other settings, evaluate our existing approach [106], and to explore two-way communication between the human and the agent (e.g. dialogue).

In essence, our agent possesses the ability to maintain the mental model of users with regards to short and long-term intentions that we can interrogate at any point in time using our 'white-box' approach. 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.

8.5 Conclusion

This chapter provides a first account of utilising a gaze-aware artificial agent in real-time for assistance in the context of strategic gameplay, which demonstrates the potential for building ubiquitous intention-aware systems. We demonstrated that we could design the communication and interaction capabilities for socially interactive agents by addressing several prerequisites for effective human-agent collaboration [149, 155]. 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.

In the previous chapter, 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 this chapter, 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.

The chapter presents two primary contributions to the design of artificial agents that collaborate with humans, from both sides of the interaction. From the agent end, we show that it is possible to develop agents that can not only predict intentions through gaze but communicate and reason about them as well. On the other end, we show that the human counterpart can be supported by a proactive agent that can only understand intentions but also communicate them through verbal means (e.g. written language), which helps maintains situation awareness while reducing visual distraction when compared to using a live gaze visualisation approach.

Overall, this chapter highlights an opportunity for nonverbal behavioural inputs to increase the human-agent interaction and provide a viable 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 as 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.

The best way to predict the future is to invent it.

– Alan Kay

9

General Discussion

9.1 Introduction

In this final chapter, we first summarise the overarching motivations of the thesis, the relationship between each study, and how they align with *human-centred prerequisites* outlined in Chapter 1. We then revisit and answer the research questions articulated in Chapter 2 (Section 9.2). The chapter then discusses the key contributions and implications of our findings arising from studies performed in this thesis (Section 9.3). This section is then followed by the limitations and future work (Section 9.4), and the concluding remarks (Section 9.5) of the thesis.

This thesis begins with a vision of the future where humans and embodied artificial agents form a 'symbiotic relationship' in all aspects of life—a future where artificial agents possess the ability to interact with humans as 'naturally' as humans interact with other humans. Recent strides in verbal communication with computers such as intelligent voice assistants are evidence that

natural communication is an important factor in forming this symbiotic relationship [199].

Humans interact with other using a full range of communication methods. The use of nonverbal communication is common in everyday human communication [138, 171]. Its inherent use begins even before infants start to develop speech or understand words. However, most forms of nonverbal communication remain difficult for computers to understand and therefore utilise.

On the other hand, we are only beginning to improve how we can interact with computers using verbal communication in recent years, but its usage remains limited (primarily due to the explicit *input-process-output* pattern). Therefore, in the thesis, we argue that for artificial agents to be able to interact as effectively as humans, they must first be able to recognise nonverbal communication cues—by knowing when and what to look out for, and by understanding the associations between them. Nonverbal communication has the potential to augment the mature developments in verbal communication and to change the way humans and computers interact with one another. Early work in human-computer interaction such as Bolt [33]'s 'Put-That-There' exemplifies the strengths and vast potential in the joint use of verbal (speech) and nonverbal (gesture) communication, which offered an effortless way to interact with intelligent systems.

In recent decades, the advancements technologies that sense human input and its respective detection algorithms have provided HCI researchers with the opportunity to explore the *utilisation of nonverbal communication* prerequisite from a computing perspective, which is a crucial prerequisite for enabling effective human-agent collaboration [149]. As such, this thesis aims to explore this prerequisite as its primary aim, focusing on the implementation of gaze as a viable input modality for enabling effective collaboration afforded by the ubiquity of eye tracking technology. From literature, we know that gaze is a rich nonverbal communication signal present in face-to-face human-human interaction. For this thesis, we combine three principles of gaze to improve the capacity of artificial agents to interact with its human counterparts: (1) *gaze signals interest*, as the human gaze is drawn to what interests them, (2) *gaze precedes action*, as humans look ahead of before performing an action, and (3) *gaze is social*, as it serves as an important signal (sender) and a cue (receiver). In addition, the thesis further explores three other human-centred prerequisites, which are also crucial components for our final study (Chapter 8).

To illustrate how gaze awareness can support human-agent collaboration, we situate the thesis in the context of a competitive gameplay setting where humans and agents form hybrid teams.

We designed a series of human-centred studies, where each informs the design and development of a collaborative intention-aware artificial agent that recognises human intention afforded by combining gaze input into classical models of intention recognition (e.g. observing actions).

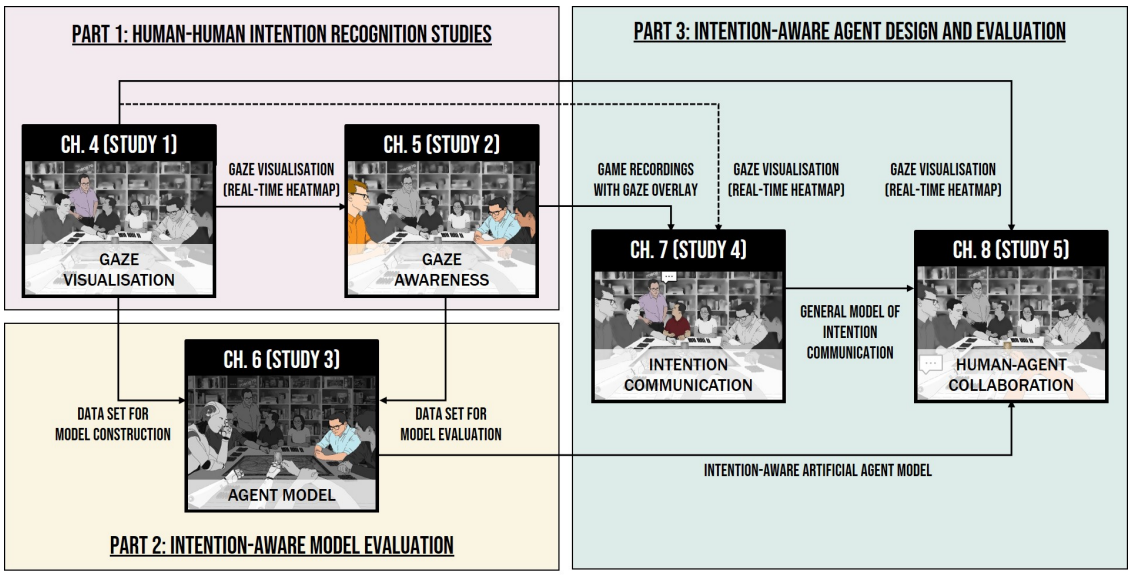


Figure 9.1: Overview of thesis studies and outcomes (and how they related to each other).

Figure 9.1 shows the progression of the studies and the relationship between each of them, i.e. how the outcomes of each study inform the next. Each study investigates the interactions between a different pair of actors in the scenario, whereby the actors include human non-player (spectator), human players or artificial agents. Study 1 and 2 investigates the utility of gaze in human-human interaction, focusing on the effects of visualisation and awareness in our setting. Study 3 evaluates the performance of the agent, given its ability to combine gaze into its planning process. Study 4 focuses on making the ability to predict intentions useful by developing the communication protocols in order for the agent to communicate its inferences as a first step. Study 5 combines the outcomes from the previous studies, particularly the evaluated prototype artificial agent in Study 3 and the communication protocols in Study 4.

Study 5 focuses on the interaction between the artificial agent and assisted-players, given the ability to predict and communicate intentions of an opponent on behalf of the human-agent team. The study, therefore, determines whether the agent can recognise and communicate intentions in real-time while evaluating the experience of the assisted players that receive the predictions about their opponent's intentions from the agent.

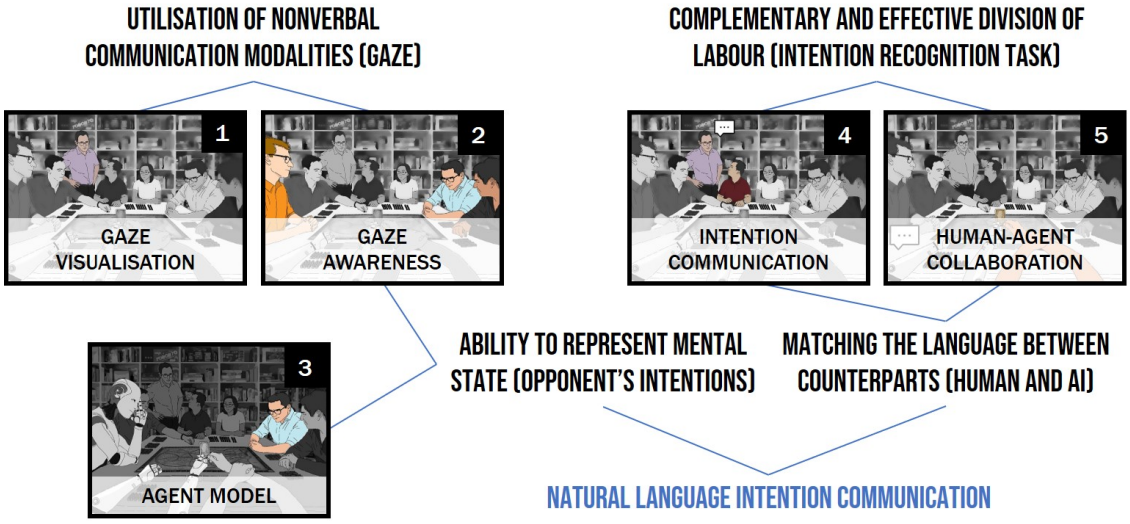


Figure 9.2: Relationship between *human-centred prerequisites* for man-computer symbiosis [155] and the studies conducted in this thesis (numbered in the top right cell).

As shown in Figure 9.2 above, Study 1 and 2 investigates the utility of gaze from a human perspective, finding its ability to predict intention through visualisation of gaze but is also susceptible to deception. As gaze acts a 'window into the mind' [122, 221], players that were aware of their opponent's gaze in Study 2 kept track of their opponent's intentions, either by recalling where the opponent has looked on the map or by using the aid of a digital map of *Ticket to Ride* they annotated on. Subsequently, this awareness allowed the players to build a mental model of their opponent's plans, and when combined with in-game actions, players were able to make predictions with the knowledge that gaze precedes actions. However, as indicated in our studies, the task of intention recognition is difficult for humans as gaze, and by extension, intentions can shift very quickly, presenting new information in a continuous stream. An observer who wishes to keep track of their opponent's plans would need to continuously update their mental model by determining which intentions are more probable. Therefore, the task of gaze-based intention recognition for humans a cognitively demanding process over time. Moreover, as shown in Study 2, human observers can be subject to biases in their observations or in their belief system, and therefore susceptible to deceptive behaviours.

Study 3 shows that the agent can effectively model the mental state of the opponent, projecting not only distal intentions but also alternative plans when gaze is combined into its planning process. As a result, the study highlights that it is possible to delegate the cognitively demanding

task of gaze-based intention recognition to the agent, enabling the player to focus on strategy formulation. Unlike automation, the task of intention recognition is complementary, as the human player also engages in the task by evaluating the state of the game. When the agent makes the same predictions as the player, it enforces the player's confidence with their own predictions. Further, the agent demonstrates the ability to provide early predictions to the player as well as an improved projection, which enables the player to be better aware of future states. The combination of these capabilities, therefore enables players to be better aware of the situation to improve decision making. Study 4 focuses on the communication of intentions inferred by the agent, allowing the assisted player to be updated about the most likely plans (awareness of future states) of the opponent through natural language. In essence, the agent has to transfer its mental model of the opponent's plans in a manageable and effective manner.

The thesis shows that by considering Licklider's four human-centred prerequisites, we developed a collaborative artificial agent that can predict intentions with the capacity to communicate them as well. From an agent perspective, Studies 1-4 resulted in two models, an 'intention-aware' artificial agent model and a preliminary language model of intention communication. As shown in Figure 3.9, the combination of these models resulted in *natural language intention communication*, in which the agent can use to both communicate and explain its inference of the opponent intentions (or plans) in natural language to the collaborative player. In turn, the ability to communicate about intentions using natural language supports the human-agent team, as it emulates to some degree how a human collaborator would perform the same task in the same setting, including the ability to explain its reasoning process (e.g. what was observed).

Ultimately, by addressing the human-centred prerequisites for the design of a collaborative agent through a series of human-centred studies, we demonstrated the potential for an agent to become a better collaborative partner. Study 5 evaluates this potential by evaluating how the agent would fare in real-time human-agent collaboration. One on end, its ability to perform gaze-based intention recognition in real-time and whether its resultant predictions are useful to the assisted-player compared to using gaze visualisation. One the other end, we evaluate how the preliminary language model fares, in terms of being able to communicate the intentions and provide transparency about its reasoning process. The study resulted in considerations for building collaborative intention-aware human-centred AI systems that incorporate gaze input.

9.2 Research Questions Revisited

This thesis investigates the use of gaze to support collaboration between human and artificial agents. To scope the thesis, we have chosen to explore how gaze awareness can support collaboration in terms of intention recognition. Knowing the intentions of collaborators and others such as opponents provides awareness of the situation. In this section, we revisit the research questions formed in Chapter 2 and summarise the findings from the corresponding studies for each question. The overarching research question addressed in this thesis is:

> **Main RQ:** How might gaze-based intention recognition improve real-time human-agent collaboration in a competitive gameplay setting?

To help answer this question, we formed three sub-research questions—RQ1, RQ2 and RQ3. As a first step, the thesis focuses on understanding the relationship between gaze and intention from a human-human perspective in our setting, forming the first sub-research, RQ1:

> **RQ1:** Does gaze awareness improve the capacity of a human player to perform intention recognition in a competitive gameplay setting?
> **RQ1.1:** How do we represent gaze to improve a human's capacity to perform intention recognition in a competitive gameplay setting?
> **RQ1.2:** What are the effects of gaze awareness for intention recognition in a competitive gameplay setting?

To answer RQ1, we focus on two aspects, the visual representation of gaze for intention recognition, as gaze visualisation is highly dependent on context and individual preference (RQ1.1), and second its effects for intention recognition in a competitive gameplay setting (RQ1.2).

Study 1 evaluated that the *real-time heatmap visualisation* had ideal characteristics that enable humans to improve their capacity to perform intention recognition, which answers RQ1.1. Study 1 informs the visualisation employed in Study 2, where gaze is displayed to 'gaze aware' players throughout the game. In Study 2, we found that gaze awareness improves the capacity of human players to perform intention recognition in a competitive gameplay setting. The players were able to make early predictions about their opponent's intentions and better triage their strategies with knowledge of their opponent's plans. However, there are several downsides despite the

ability of gaze awareness to improve intention recognition, which led to no net gain in win rates. The adverse effects of gaze awareness include the increased cognitive load due to the added information and the inability to attend correctly to the task at hand. In general, players with increased awareness adopted non-optimal strategies (in some sense aggressive, e.g. more blocking) that were detrimental to their own plans when they became aware of their opponent's.

With a better understanding of the relationship between gaze and intention, we shifted our attention towards improving the intention recognition capacity of a prototype artificial agent by incorporating their gaze into its planning process. As part of this thesis, we have begun the development of a prototype agent alongside from its inception. Using the labelled data set from Study 2, we evaluated its performance accordingly, forming sub-research question RQ2.

> **RQ2:** Does gaze awareness improve the capacity of an artificial agent to perform intention recognition?

The artificial agent uses gaze input as *priors* and probability distributions to form potential plans. In short, areas of the interface (i.e. game board) that have received more visual attention will result in higher probabilities, and the agent will generate potential plans according to the distribution and in-game actions that have been played. The evaluation in Study 3 showed that gaze improves the capacity of the prototype artificial agent to perform intention recognition. Further, the agent performs better and is less prone to deceptive behaviours than the human subjects in the human-human intention recognition study (Study 2). We also demonstrated that the agent works well for predicting long-term intentions and that it can start to make early inferences when no in-game actions have been played. However, the agent does not perform well in situations where no plans or only short-term plans (quickly changing) have been formed.

Nevertheless, the results demonstrate that gaze improves the intention recognition capacity of artificial agents, answering RQ2. The third and final sub-research question, RQ3, therefore poses a broader question with regards to the collaborative design aspects of gaze-enabled agents.

> **RQ3:** How do we design gaze-based intention recognition systems for supporting human-agent collaboration?

To answer RQ3, we focused on two aspects, the communication capabilities of the agent and whether its adoption is beneficial in our scenario, forming RQ3.1 and RQ3.2, respectively.

> **RQ3.1:** How should an artificial agent communicate intentions given the ability to perform gaze-based intention recognition?
> **RQ3.2:** Does delegating the task of intention recognition to the agent improve the player's decision making in a competitive gameplay setting?

To answer RQ3.1, we designed a human-centred study that elicits how human collaborators would communicate in an assistive fashion, given the ability to perform gaze-based intention recognition using gaze visualisation (Study 4). The study produced a preliminary *language model of intention communication* that provides the format in which predictions should be structured for effective communication, along with the content properties to include in an explanation when it forms parts of the 'message' that is received by an assisted human collaborator.

Study 5 demonstrates the use of the agent in our chosen competitive gameplay setting, where the agent assists an active human player. The agent autonomously makes predictions about the likely plans of the team's opponent as a human teammate would, enabling it to interact with a human collaborator proactively. The fact that we were able to assign the role of gaze-based intention recognition to the agent is evidence of its collaborative capacity. Study 5, therefore, evaluates whether the adoption of the agent together with the language model improves the decision making of assisted-players. In theory, the increased 'situation awareness' of the agent combined with the reduced perceived cognitive load from the delegation of the cognitively demanding task of gaze-based intention recognition would improve the decision-making ability of players.

Study 5 finds that the delegation of the intention recognition task only improves the decision making of players to some extent. Although the assisted players reported a positive overall experience when engaging with the agent, the study revealed several further design considerations for human-agent collaboration to be effective. These considerations include the representation of intentions from the player's perspective and to determine what the assisted-player already knows or intends to do from the agent's perspective, which opens up new avenues for future research. Returning to the Main RQ, the combined findings from the sub-research questions demonstrate that gaze-based intention recognition has the potential to improve real-time human-

agent collaboration in competitive gameplay settings. This demonstration is made possible from the positive findings in the thesis but requires further exploration. In the next section, we will further unpack the implications arising from the contributions from answering the RQs.

9.3 Key Contributions and Implications

9.3.1 Extending the Relationship Between Gaze Awareness and Intention

The first contribution of the extends the knowledge surrounding the relationship between gaze awareness and intention. We contribute the knowledge that observing the gaze of human subjects over time allows humans and agents alike to recognise their long-term plans (distal intentions). This capacity is facilitated by the ability to record and process eye movement data in real-time, along with developments in the field of AI. The literature so far has only captured short-term intentions from observing gaze, mainly due to the task and scenario chosen (e.g. sandwich making). Unlike recognising short-term intentions (e.g. what someone might do next), being able to recognise distal intentions provides a bigger picture of the situation and provides the opportunity to act upon the information. This contribution has direct implications for designing how human could interact with collaborative agents in the future. Moreover, by enabling gaze-supported 'intention awareness' in artificial agents, it complements other forms of awareness in future agents, such as cognition- and context awareness [3, 44], and therefore, forms part of the puzzle in endowing virtual agents or embodied agents (e.g. robots) with human-like abilities.

The thesis further demonstrates that gaze can be employed deceptive ways (together with actions) when human subjects become aware that their gaze is being observed by others, particularly in adversarial settings such as the scenario employed in this thesis. Chapter 5 provides a *taxonomy of gaze-based deception strategies* elicited from participants that attempted to obscure their gaze to avoid their intentions from being known. In the future, it is possible that humans with the knowledge that intelligent computer systems are continually tracking their nonverbal cues, and may attempt to obscure their behaviours. Hence, for gaze input at least, we can use the taxonomy together with the data set on gaze behaviours, we can begin to investigate how systems can disentangle between natural and deliberate gaze behaviours, and to design intelligent gaze-enabled systems that are robust to deception or when gaze is 'noisy' or even partial.

The preliminary analysis in the addendum to Chapter 5 shows that gaze behaviour do differ between natural gaze behaviour and deceptive conditions in terms of features. At the same time, the results in Chapter 6 suggests that the artificial agent is less susceptible to deception than human subjects. Together, it seems plausible to build gaze-aware AI systems that are not only robust to deception but able to detect deception as well. Hence, we propose a two-pronged approach for this implication. On one end, we can train the agent to determine how natural gaze behaviours look like, either independent or dependent to activity using eye movement sets. Any difference from this model can raise flags on deceptive behaviour (i.e. deviation from a baseline). On the other end, we can use the taxonomy as a starting point to model how deception looks like using classification techniques. Then again, the data set from Study 2 is to some degree limited as it does not distinguish the points in time when players employed deceptive behaviours.

To address this limitation, future research into gaze-based deception could begin by collecting an empirical data set based on the strategies presented in the taxonomy. Participants in a well designed study can employ a 'deceptive frame of mind' and exhibit the corresponding behaviours purposefully, eliciting unique gaze behaviours when compared to non-deceptive use. This analysis would result in complex process, requiring researchers to build a dictionary of deceptive behaviours for each strategy starting from the lowest level of features. For example, for *gaze decoy*, we can start to trace the features when they are looking at their own plans versus a decoy path with ground truths. Behaviours such as staring deliberately at an area or object for a prolonged period of time may signal that a decoy plan, as it would be considered 'unnatural' to have a long fixation duration but low fixation count. At the same time, the addition of saccadic features provides additional insights into the different areas of interest (i.e. areas of actual plans versus areas of decoy plans), as the deceiver has to look between them periodically. Using the difference in features, we can then start to model the patterns of behaviours for each strategy.

Beyond the vast implication for considering the use of gaze-based intention recognition in AI systems, especially for collaborative settings, this findings on gaze-based deception have direct implications for non-invasive multimodal deception detection [2, 46]. As gaze provides early prediction on intentions, it can further be combined with existing approaches for early detection of deceptive intentions. The existing approaches includes *physiological* (e.g. thermal imaging), *visual* (e.g. facial expressions and gestures), and *speech* (e.g. verbal style and content) data.

However, our findings so far showed that detecting the differences in gaze behaviour is difficult in line with existing literature (e.g. [152, 205]). Even with non-deceptive natural gaze and in-game behaviours, the plans of players, for example, may change quickly, making it challenging to disentangle the intentions that are abandoned as well. Moreover, as humans can simultaneously possess multiple intentions, it increases the difficulty in predicting which of the intentions is likely to be carried out next. In such cases, our approach of combining multiple intentions together in a plan with alternatives considered (e.g. top 10 plans) may provide the human-agent team's insights into the goals of the human, instead of just considering the most likely plan.

9.3.2 Design of a Collaborative Intention-Aware Artificial Agent

The second contribution of the thesis is the design and initial evaluation of a *collaborative gaze-enabled intention-aware agent*, which constitutes an artefact contribution. In this thesis, we constructed the agent from the ground up by carefully designing human-centred studies that inform the different components of the agent, including the collection of a data set for its evaluation. In essence, we tackled the two core components required for the design of a collaborative artificial agent: the automated recognition of human intentions using gaze (Chapter 6) and its communication protocols once recognised (Chapter 7). We further considered the explainable capabilities of the agent—an important aspect for building trust in human-agent collaboration. As a result, the artificial agent is not only capable of inferring intentions using the combination of implicit gaze and explicit in-game actions but also capable in explaining how it came about its inferences using natural language to a human collaborator.

The design of the preliminary language model in which we considered both the communication of intentions and its corresponding explanation based on what the agent 'sees' has direct implications for the use of nonverbal communication in future Explainable AI (XAI) systems. Study 4 showed that humans have a unique but natural way to abstract information when explaining their reasoning process, especially given the ability to make predictions using gaze. Hence, an artificial agent should possess the same capability to do so in collaborative settings. The language model is based on how humans communicate and subsequently explain the intentions of others. The model that takes into account different levels of information, from low-level observations to higher-level associations (e.g. between different areas or objects, between explicit past actions

and implicit gaze). These behaviours can then be mapped onto text representations in natural language which essentially abstract the information, enabling the assisted player to receive the most likely intended plans of the opponent during each communication to prevent information overload and to reduce the frequency of communication (as to not interrupt the task at hand).

Chapter 8 combines the outcomes of the previous chapters to further evaluate the design of agent while focusing on the experience of players as they played alongside the agent. The study provided further design implications with regards to information presentation, the need for context awareness, usage of nonverbal signals in human-agent collaboration, and the agent's explainable agency. The thesis also demonstrates that *human-centred design* approaches can be employed for designing and evaluating collaborative multimodal artificial agents of tomorrow. As such, the thesis exemplifies and reinforces how human-computer interaction and interaction design can better inform the design of human-centred AI systems.

9.3.3 Demonstration of Gaze Awareness for Human-Agent Collaboration

The third contribution of the thesis is the demonstration of gaze awareness for human-agent collaboration. The working prototype artificial agent highlights the importance of incorporating gaze to improve the inference of human intentions in real-time through nonverbal communication. By incorporating gaze input into the planning process of a 'white-box' model-based artificial agent, it effectively extends the interactive capabilities of the artificial agent in three ways:

1. Gaze awareness improves the agent's capacity to predict potential future states much earlier and more accurately (i.e. long-term/distal intentions). This allows a human in collaboration to adjust their plans accordingly to suit their goals, or confirm their own predictions, and therefore improve their ability to make better-informed decisions.
2. Gaze awareness allows the agent to initiate the interaction by providing timely situation awareness updates (or queries) proactively. This consequently narrows the "interaction gap" between the two entities beyond the ability to recognise human nonverbal cues. This ability is especially beneficial when the agent is aware of the intentions of both the collaborator and of others, which would allow the agent to provide contextually-relevant information or recommendations that directly impact the goals of the human collaborator.
3. Gaze awareness allows for the division of labour in human-agent collaboration. In our

work, we gave the delegated the task of intention recognition to the artificial agent to reduce the human collaborator's perceived cognitive load while maintaining their situation awareness. This division of labour, in turn, allows the human collaborator to better focus on their tasks and decision making (Chapter 8). Given the agent's potential to act autonomously, there are a multitude of ways the agent can provide assistance, for example, by taking over tasks of the user if the user is currently overloaded based on their gaze.

However, we note that were able to demonstrate these capabilities as we factored Licklider's *human-centred prerequisites* as part of the considerations of the agent. As such, the combination of the artificial agent model (that effectively keeps track of a user's mental model), the preliminary language model for communicating and explaining inferred intentions in natural language, and the delegation of the distracting and cognitively demanding task of intention recognition to the artificial agent presents an ideal situation for the thesis. This combination provides a scenario where automated gaze-based intention recognition in collaborative settings can prove to be beneficial while highlighting the importance of addressing the multiple human-agent collaboration prerequisites to improve the overall experience of the human collaborator.

9.3.4 Implications for Gaze in Human-Agent Collaboration

In this thesis, we posit that 'for artificial agents to work with their human collaborators effectively, artificial agents first need to harness nonverbal communication cues commonly present in human-human interaction'. Towards this statement, the thesis worked towards incorporating gaze into a prototype artificial agent, which allowed the agent to increase its capacity to perform intention recognition. In this thesis, we demonstrated how artificial agents could become better collaborative partners by giving them the ability to infer the intentions of human partners through their eye gaze, which have direct implications for *gaze in human-agent collaboration.*

The thesis adopts the same shift in thinking where computers should be viewed as partners rather than as tools. One of the aims of the thesis is to demonstrate that an artificial agent with gaze awareness can potentially become an effective collaborative partner to change the perception of everyday human users. As we took into account the multiple prerequisites for human-agent collaboration when designing the interactions with the agent, the demonstration of its collaborative capacity is evidence of its importance. Prior work has already demonstrated

the importance of using other nonverbal communication cues such as gestures [149], while in this work, we contribute the use of gaze into the mix. In the future, the combination of these cues with the ability to make sense of them will open up new avenues for the field of human-agent collaboration—and for Licklider's *Man-Computer Symbiosis* [155] vision to be realised.

In Chapter 2, we described how collaboration is not the same as interaction, as they involve a subset of characteristics that all parties need to possess for effective collaboration (e.g. mutual goal understanding, preemptive task management and shared progress tracking [29]). Humans can collaborate with minimal effort as they use a range of communication cues, which helps mitigate communication overhead on a single channel or necessary communication. An artificial agent that is aware of the situation, can reduce communication overhead by only interacting when required. At the same time, the agent can also determine *when* and *how* it should interact. Moreover, we can look towards interacting with embodied artificial agents, such as in intelligent social robots. Our work demonstrates the potential to give robots the ability to infer human intent. Therefore, we can start to imagine how robots might collaborate with humans in everyday life. For example, can a robot prepare to hand me a tool that I need from inferring my intentions?

One of the contributions of this thesis is the capacity of artificial agents to recognise 'long-term' (or distal) intention recognition, in which we also proposed a generalisable approach for implementation. Therefore, we can also begin to explore its use in complex human-agent teaming settings where knowing the intentions of others is essential (e.g. air traffic control, wargaming).

9.3.5 Implications for Gaze in Human-AI Interaction

Based on current predictions, it is likely that most Human-Computer Interaction will eventually shift towards Human-AI Interaction. Artificial intelligence in recent years has transformed changed how we interact with technology and will continue to do so [260]. One of the broad implications of our work is how AI can be used to mediate gaze input in human-human interaction.

One of the primary motivations for adopting an artificial agent to process gaze is the adverse effects of using gaze visualisation. Gaze visualisation presents a straightforward way to implement gaze input to create 'awareness' in human-computer interaction, particularly in computer-mediated interaction. This approach hinges on the ability of the human observer to make sense of the information. As gaze is highly correlated with the locus of attention [24], it has been

shown to be beneficial in most scenarios. However, there are limits with regards to its use, especially when we start to think about *scalability*. For instance, it might be possible to scale gaze up to three users in a triadic collaboration scenario using gaze visualisation. However, beyond this number of users, it will begin to clutter the interface. Hence, our approach of using an agent to infer gaze poses a solution to the problem of scalability. The gaze-enabled agent can act as a mediator between users or human teams, one that can infer the intentions of all users and make personalised recommendations as well based on the intentions of all parties.

Remote work has become an important theme in 2020, and prior work has shown the strengths of employing eye tracking in remote collaborative settings into computer-meditated communication (CMC) scenarios (e.g. [11, 143]). Through this work, we can start to see how gaze-enabled AI can further improve collaboration through mediation between humans as a starting point, such as abstracting intentions of large teams to find common ground. Further, we can look towards other domains such as learning, where multimodal sensing is being explored [89, 229, 230, 228].

Our work has implications towards novel paradigms for interaction with automated systems, such as Schmidt and Herrmann [218]'s *Intervention User Interfaces*. By knowing the intentions of a user and of others (especially implicitly), the system can not only adapt to the situation but also *intervene* when required, such as in to prevent errors or the warn users. More broadly, the contributions in this thesis highlight the importance of incorporating nonverbal communication in human-AI interaction. In existing work, nonverbal modalities such as gaze have been used for context- and cognition-aware ubiquitous interactive systems [45]. The ability to recognise intentions (i.e. 'intention-awareness') shown adds another dimension to these forms of awareness. To elaborate, context-awareness allows the AI to know what the user is currently doing, which would allow them to adapt their interaction according, such as to provide contextually-relevant information, reflective of how humans interact and communicate with others. Subsequently, cognition-awareness supports information intake, which allows the AI to determine when it is the right time to convey the information based on the user's cognitive load and engagement, and further optimise the level of communication required. Lastly, intention-awareness allows the AI to be aware of what users may be planning in the short- and long-term. This capability enables AI systems to establish common ground while maintaining the autonomy to provide crucial information or perform actions ahead of time, based on the inferred intentions of all users.

9.4 Limitations and Future Work

9.4.1 Artificial Agent Intention Recognition Performance

In this thesis, we developed and evaluated a prototype 'intention-aware' artificial agent model that incorporates gaze into its planning process, showing that gaze input increases its capacity to infer intentions. We opted for a model-based approach to intention recognition as there we did not have sufficient data to train a data-driven model, but also that it allows to us interrogate the model (i.e. 'white-box') and therefore affords higher explainability. An inherent limitation of using a model-driven AI approach for processing gaze data is that it is difficult to model the different gaze behaviours that may represent intentions, resulting in lower accuracy. However, this differentiates our approach with the existing work which adopted data-driven approaches (Section 2.5), and that it has the potential to be generalised with knowledge of the domain.

The prototype implements two fixation-based features—*fixation duration* and *fixation count*—to update its gaze model. Using these basic 'low-level' gaze features, we increased the weights of the cities in two ways: when the player fixates on the cities (nodes) and on the routes that are related to the cities (edges). As an example, the probability for Denver also increases when the player fixates on the route connecting Denver to Helena. The two features allowed for the agent to handle situations where players looked at an area for a long period and repeatedly looked at an area of interest. While this weighting method provides an acceptable way to generate plans based on interest based on our evaluation, the agent does not take into account the patterns of gaze behaviour. For instance, when a tracked player constantly compares two areas of interest, a human observer can easily discern such patterns and factor them into their strategy. A gaze-enabled agent should be able to make sense of the same observations as well.

In future work, we can potentially combine our model-driven approach with a data-driven approach to increase the overall performance of the agent, in addition to improving the model itself. As a first step, we can potentially increase the agent's intention recognition performance by incorporating the knowledge abstracted from analysing the gaze patterns of a tracked user. The use of gaze patterns in an intelligent interactive system is well demonstrated in Qvarfordt and Zhai [202]'s *RealTourist*. For instance, their study showed that gaze patterns could reveal when a user is no longer interested in a particular object, which has direct implications for our

artificial agent. The current agent iteration lacks a decay function to manage its probability distribution, whereby the areas that have not received gaze after some time has passed could signal that the user is no longer interested. Without such a function, the agent considers all the detected points of interest from the start of the game even when plans change.

The consideration of multiple generated plans (e.g. top 10 most likely plans) instead of only considering the top most mitigates this limitation to some extent. Hence, the ability to detect disinterest through understanding gaze patterns would allow the agent to improve its performance by disregarding the intentions that have are no longer relevant. The *RealTourist* paper identifies two ways this can be detected, either using a temporal approach such as a decay function (i.e. reducing probability of the objects that no longer receives fixations after some time) or by understanding gaze pattern that signal disinterest i.e. a data-driven approach [201]. Additionally, the inclusion of gaze patterns in future work can improve the automated mapping between the detected gaze behaviours and its corresponding explanation. As a result, the agent can further improve how it explains its predictions by discerning what a tracked user is trying to do, reflective of the inferences using the rule-based assembly by the 'wizards' in Chapter 8.

9.4.2 Artificial Agent Communication Aspects

Information Presentation

A significant limitation of our approach in Study 5 is the full use of textual representations to convey human intentions (Chapter 8). While this serves as a good starting point, it caused the players 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 to understand the predictions (i.e. the location of cities or map areas). Further, the text-based updates essentially took the assisted players 'out of the game', as they had to read and understand the predictions communicated before acting on them. It is also possible that players were forced to divide their attention between the game and the predictions that appeared in the pop-up. These limitations suggest the manner in which the agent communicates intentions requires further exploration.

In future work, we can potentially address this limitation by either exploring novel techniques for presenting the information to a user, or by allowing the user to interact with the information

but in a way that it is well-integrated into the interface. In our case, we can use a hybrid approach by combining a visual overlay over the interface (i.e. map) to highlight the parts of the map with precise intentions and augmented the information with on-screen natural language annotations that include explanations (see Figure 9.3). We call this *visual explanation of intentions*. A step further would be to enrich the combination with other modes of output such as speech for narration or to provide multimedia clips to provides insights. These approaches are akin to a co-located human teammate pointing (using a finger) at the screen while providing explanations.

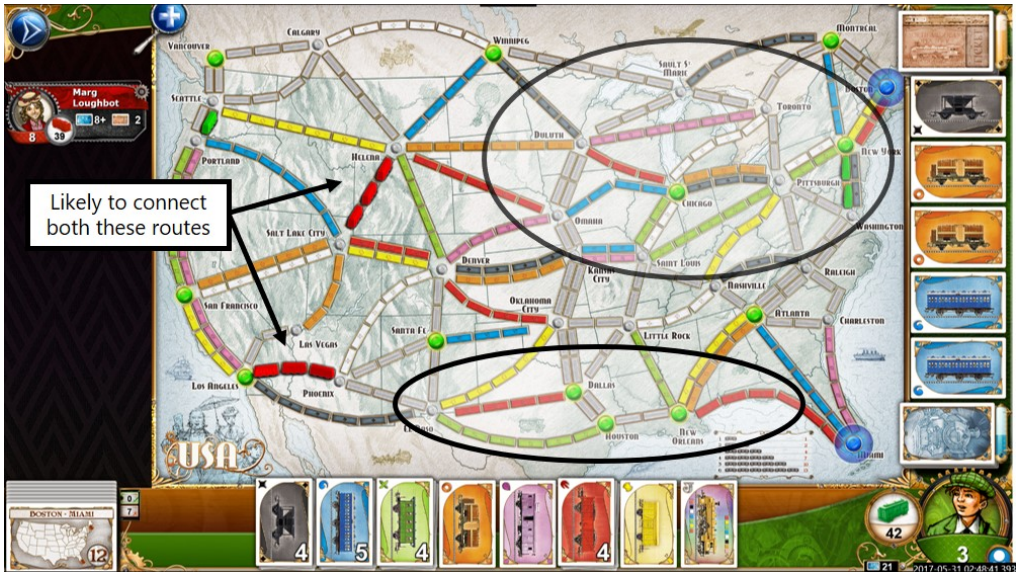


Figure 9.3: Example of visual explanation over the Ticket to Ride game.

An alternative way to present the information from gaze is the use of non-live gaze visualisation, as mentioned in Chapter 5. Our findings in Chapter 8 suggests that using a visual overlay of gaze does have some merits, such as enabling players to quickly to understand the information without knowledge of the interface (map of *Ticket to Ride*). We acknowledge that some information that players might find critical may be lost when moving from gaze visualisation to using the AI for abstraction. Non-live gaze visualisations could also help to avoid the distraction effects observed in our studies. Therefore, we can potentially design and implement mechanisms that give each player a summary of where their opponent has looked, and for how long. A similar implementation is *GeoGazemarks*, where fixations were recorded and clustered to give a history of a user's points of interest to support free map exploration [90]. Chapter 5 additionally suggests that it would be beneficial to give players some control over the gaze visualisation style, such as the ability to change its appearance or switch it on and off to manage the flow of information.

One of the limitations in Study 5 is that the players had no control over *when* the predictions were presented, causing players to be interrupted with the pop-up or distracted by the new information provided. A pop-up notification without showing predictions outright may allow for an improved experience, enabling the user to access the information if needed or when convenient (e.g. not while they are taking their turn). These capabilities can be integrated by providing the agent with the knowledge that the user is currently engaged in an activity (context-awareness). The agent can further manage the cognitive load of the user by holding back new information, scaling back the information or taking over a task for the user (cognition-awareness). Hence, in future work, we will further explore the *when* and *how* aspects of presenting the information, and the effects of giving users control over how they access and interpret the information.

One-Way Communication

In our final study, Study 5, we operationalise the agent to infer intentions of the opponent on behalf of an assisted player (Chapter 8). The agent, therefore, abstracts the intentions from its observations (gaze and in-game actions) and subsequently communicates them periodically in natural language. This development trajectory adheres to our goal of providing an artificial agent with the capability of performing gaze-based intention recognition, as stated in Chapter 1. This capability is then combined the preliminary language model of intention communication derived in Chapter 7 for the agent to provide support to a human collaborator. Consequently, the agent's primary mode of collaboration is solely to support the player by continuously predict the intentions of the team's opponent and subsequently communicating them.

On the other hand, the human player's role is to focus on playing the game, as the focus of the study was to investigate the effects of using an intention-aware gaze-enabled artificial agent. As described in Chapter 8, this interaction approach is akin to 'AI-advised human decision making', where the system provides information and the human makes the final decision based on the information past and present [22]. Bansal et al. [22]'s motivation for approach comes from the fact that humans and agents have complementary strengths and abilities, which mirrors Licklider [155]'s *Man-Computer Symbiosis* vision. Some similar examples presented are interactions with machine-learned models that are not perfect but are helpful such as for medical diagnosis or object classification. With limited ability to provide information to the agent, it limits the

support the agent can provide, such as the ability to engage in collaborative decision making. Furthermore, when we look novel paradigms for interaction such *Intervention User Interfaces* [218], our artificial agent being autonomous would also need to know when to *intervene*.

In future work, we will investigate the effects of two-way communication between the human and the artificial agent. The agent, with its ability to perform gaze-based intentions recognition, presents novels ways of interaction when we start to think about what the types of queries that the agent can potentially handle. One approach is to allow the human to naturally converse with the agent informed by using dialogue models (e.g. [159]), which has several positive effects on the interaction between the human-agent team. For instance, the agent with the ability to consider additional input from the player such as 'what-if' queries will allow the agent to generate alternative predictions based on intentions to support the human player in decision making. Moreover, conversing the agent can help the human-agent team enables the team to find common ground, such as through negotiation. This capability is crucial when their mental models of the opponent's intentions do not match, such as when the human player does not agree with the predictions made by the agent or is not clear about the predictions that were made. An agent with the capacity to factor in various input from the human collaborator can further enable the team to engage in more complex collaborative activities, and beyond the screen.

Further, in Chapter 8, the agent provided predictions at two levels, *abstract* (prediction only) and *detailed* (prediction with explanation). This is reflective of the levels in the Situation Awareness-based Agent Transparency (SAT) model [58] adopted, where abstract predictions are *Projections* (Level 3), and detailed predictions are *Projections+Reasoning* (Levels 2 and 3). Given the ability to converse with the agent, it essentially provides the human collaborator with the ability demand for the information instead of the agent providing periodically. This 'demand-driven' approach for requesting information as needed has been explored in preliminary study [244], where authors investigated demand-driven transparency for monitoring intelligent agents to avoid information overload. In a study with 36 participants, in which participants were given control over the level of transparency level ('demand-driven') derived from the levels in SAT model, the results indicate an improvement in performance, response time, usability, trust and error detection. Hence, in future work, with two-way communication can give users this control.

However, we must be wary that two-way communication is more time-consuming because the

human user has the opportunity to feedback and to question what has the agent has communicated. In our case, it might take the player 'out of the game' as their attention is focused on what information the agent can provide. Using the experimental setup in Study 5, Madumal et al. conducted a study to validate an existing model of discourse-driven explanation (two-way communication) [160]. In the study, players communicate with the 'wizard' through the chat application, and either the 'agent' or assisted-player can initiate the interaction. The 'agent' could provide predictions about the opponent's plans and also explain its reasoning if required. The observations of the study showed that the interactions between the players were relatively short, suggesting that the player did not engage the agent for long periods. The dialogue ends when players obtain the information they required, which was more efficient as they had control over what to query. The dialogue also ends when the agent's assistance is no longer required. Therefore, the study suggests that human collaborators with control over the information they require, such as *when* and *what* (in line with 'demand-driven transparency' [244]), using two-way communication would potentially improve the overall collaboration in the human-agent team.

9.4.3 Artificial Agent Collaboration Capacity

The thesis presents an account of humans collaborating with an artificial agent that can perform gaze-based intention recognition, in our case, collaborating against opponents in a competitive gameplay setting. The studies in the thesis focused on inferring the *intentions of an opponent* as a first step. Hence, in the final study of the thesis, Study 5, the agent supports the human collaborator by solely communicating the intentions of the opponent through natural language. This has inherent limitations with regards to the collaborative capacity of the agent. For instance, the agent does not provide contextually-aware information about the intentions of the opponent and therefore might communicate information that may be redundant to the user, or irrelevant to the plans of the player. While the agent performs the task of gaze-based intention recognition on behalf of the human-agent team, it only presents a small step towards is potential collaborative capabilities. In terms of degrees of symbiosis, the interaction represents a *commensalistic* symbiotic relationship, where one party benefits but the other is neither harmed or helped [70].

In future work, we plan to increase the artificial agent's collaborative capacity by factoring in the intentions of the assisted player as well. The combination of the intentions of both the

assisted player and the opponent, which can be inferred implicitly through gaze, will enable the agent not only provide intentions that may influence the plans of the player but also provide recommendations based on the knowledge of the intentions of both players—in some ways, working towards a *mutualism* symbiotic partnership, where both parties benefits as partners [70]. Nevertheless, recommendations from AI systems—even if imperfect—can result in human-AI teams that perform better than either the human or the AI system alone [22].

This potential for this capacity is demonstrated in Study 5, where we show that the agent can process and communicate intentions in real-time by observing a human player (i.e. the opponent) and that is possible for the agent to track the intentions of the assisted-player simultaneously. The plans of the assisted-player can also be communicated explicitly to the agent by means of dialogue, annotation, other forms of interaction. Therefore, for an artificial agent provide recommendations in our scenario, the agent needs to keep track of the mental models of both players and then discriminate them accordingly. As a first step, the graph visualisation may offer some ideas on how this might look like for the agent in real-time (see Section 6.4.4). The graph can be used to display the plans of both players from our data set simultaneously and visualise any overlap in plans in a different colour. By employing a similar human elicitation study as described in Chapter 7, we can start to determine how humans convey *recommendations* with knowledge of the plans of both players for collaboration using the graph.

Moreover, the condition in Study 5 where the agent communicates detailed predictions, was an extreme case where we gave the 'most complete' explanation possible without considering what the assisted-player already knew, leading to the communication of redundant information. This redundancy is exacerbated when the predictions of the agent did not change, as the same information would be communicated over again. 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. By doing so, it will prevent the agent from repeating the same information over again, and adapt it according to whether it needs to be communicated again (e.g. on critical information). 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, therefore improve the human and agent to collaborative with each other.

9.5 Concluding Remarks

The thesis investigates how gaze-based intention recognition supports real-time human-agent collaboration in a competitive gameplay setting which highlights the importance of incorporating nonverbal communication in human-AI interaction. The use of nonverbal communication is common in humans, especially for coordinating collaboration, but it is difficult for computers to understand and therefore utilise. Through human-centred design approaches, the thesis demonstrates the role of gaze awareness in increasing the intention recognition capacity of humans and artificial agents alike, in terms of accuracy and timeliness. As a result, the thesis contributes a working prototype agent that can not only increase its capacity to perform intention recognition with gaze, but also its ability to support a human collaborator by communicating the infer intentions in natural language. When designing the communication protocols of the agent, we further considered its capacity to explain, which is important for building trust with the agent—afforded by our 'white-box' planning approach, allowing us to interrogate the model.

Overall, the thesis highlights the importance of nonverbal communication for supporting human-agent collaboration. We broaden our insights into the relationship between gaze and intention, showing that artificial agents can be better collaborative partners with its ability to keep track of intentions. At the same time, agents with the ability to explain can better support in terms of finding common ground and mutual understanding for human-agent teams. These capabilities are reflective of how partners collaborate and by using the agent, which not only emulates the capability of human collaborators but can augment human capability as they can recognise intentions earlier and more accurately. Our success so far shows the strengths of adopting a human-centred design approach towards the design of collaborative AI systems.

Lastly, the thesis shares the same vision as Licklider's *Man-Computer Symbiosis*, where humans and computers would develop a 'symbiotic relationship', where the strengths of one counterpart would counterbalance the limitations of the other. The success of our agent is possible as we have considered the prerequisites. As discussed, there is further scope to improve the agent's capabilities, especially in terms of its performance, communication and collaborative capacity. In future work, we will explore these improvements to further evaluate the strengths of using nonverbal communication for human-agent collaboration in hopes that future interactions between humans and AI will continue to become more social, natural and effective.

Bibliography

[1] Yomna Abdelrahman, Anam Ahmad Khan, Joshua Newn, Eduardo Velloso, Sherine Ashraf Safwat, James Bailey, Andreas Bulling, Frank Vetere, and Albrecht Schmidt. "Classifying Attention Types with Thermal Imaging and Eye Tracking". In: *Proceedings ACM Interactive, Mobile, Wearable and Ubiquitous Technologies*. IMWUT 3.3 (Sept. 2019), 69:1–69:27. DOI: 10.1145/3351227.

[2] Mohamed Abouelenien, Veronica Pérez-Rosas, Rada Mihalcea, and Mihai Burzo. "Deception Detection Using a Multimodal Approach". In: *Proceedings of the 16th International Conference on Multimodal Interaction*. ICMI '14. Istanbul, Turkey: Association for Computing Machinery, 2014, pp. 58–65. ISBN: 978-1-4503-2885-2. DOI: 10.1145/2663204.2663229.

[3] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. "Towards a Better Understanding of Context and Context-Awareness". In: *Handheld and Ubiquitous Computing*. Ed. by Hans-W. Gellersen. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999, pp. 304–307. ISBN: 978-3-540-48157-7. DOI: 10.1007/3-540-48157-5_29.

[4] Henny Admoni, Bradley Hayes, David Feil-Seifer, Daniel Ullman, and Brian Scassellati. "Are you looking at me?: Perception of robot attention is mediated by gaze type and group size". In: *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE Press. 2013, pp. 389–396. DOI: 10.1109/HRI.2013.6483614.

[5] Henny Admoni and Brian Scassellati. "Robot Nonverbal Communication as an AI Problem (and Solution)". In: 2015. URL: https://www.aaai.org/ocs/index.php/FSS/FSS15/paper/view/11656.

[6] Henny Admoni and Brian Scassellati. "Social Eye Gaze in Human-robot Interaction: A Review". In: *J. Hum.-Robot Interact.* 6.1 (May 2017), pp. 25–63. ISSN: 2163-0364. DOI: 10.5898/JHRI.6.1.Admoni.

[7] Henny Admoni and Siddhartha Srinivasa. "Predicting user intent through eye gaze for shared autonomy". In: *Proceedings of the AAAI Fall Symposium Series: Shared Autonomy in Research and Practice (AAAI Fall Symposium)*. AAAI Press Toronto, ON. 2016, pp. 298–303.

[8] Kumar Akash, Tahira Reid, and Neera Jain. "Improving Human-Machine Collaboration Through Transparency-based Feedback – Part II: Control Design and Synthesis". In: *IFAC-PapersOnLine* 51.34 (2019). 2nd IFAC Conference on Cyber-Physical and Human Systems CPHS 2018, pp. 322–328. ISSN: 2405-8963. DOI: 10.1016/j.ifacol.2019.01.026.

[9] Deepak Akkil, Prasenjit Dey, Deepshika Salian, and Nitendra Rajput. "Gaze Awareness in Agent-Based Early-Childhood Learning Application". In: *Human-Computer Interaction - INTERACT 2017*. Cham: Springer International Publishing, 2017, pp. 447–466. ISBN: 978-3-319-67684-5. DOI: 10.1007/978-3-319-67684-5_28.

[10] Deepak Akkil and Poika Isokoski. "Gaze Augmentation in Egocentric Video Improves Awareness of Intention". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 1573–1584. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858127.

[11] Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. "GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks". In: *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 1151–1158. ISBN: 978-1-4503-4082-3. DOI: 10.1145/2851581.2892459.

[12] Deepak Akkil, Biju Thankachan, and Poika Isokoski. "I See What You See: Gaze Awareness in Mobile Video Collaboration". In: *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*. ETRA '18. Warsaw, Poland: Association for Computing Machinery, 2018, 32:1–32:9. ISBN: 978-1-4503-5706-7. DOI: 10.1145/3204493.3204542.

[13] James F. Allen, Curry I. Guinn, and Eric Horvtz. "Mixed-Initiative Interaction". In: *IEEE Intelligent Systems and their Applications* 14.5 (Sept. 1999), pp. 14–23. ISSN: 2374-9423. DOI: 10.1109/5254.796083.

[14] Fraser Allison, Joshua Newn, Wally Smith, Marcus Carter, and Martin Gibbs. "Frame Analysis of Voice Interaction Gameplay". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–14. ISBN: 9781450359702. DOI: 10.1145/3290605.3300623.

[15] Sean Andrist, Wesley Collier, Michael Gleicher, Bilge Mutlu, and David Shaffer. "Look together: analyzing gaze coordination with epistemic network analysis". In: *Frontiers in Psychology* 6 (2015), p. 1016. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.01016.

[16] Sean Andrist, Michael Gleicher, and Bilge Mutlu. "Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 2571–2582. ISBN: 9781450346559. DOI: 10.1145/3025453.3026033.

[17] Stavros Antifakos, Adrian Schwaninger, and Bernt Schiele. "Evaluating the Effects of Displaying Uncertainty in Context-Aware Applications". In: *UbiComp 2004: Ubiquitous Computing*. Ed. by Nigel Davies, Elizabeth D. Mynatt, and Itiro Siio. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 54–69. ISBN: 978-3-540-30119-6. DOI: 0.1007/978-3-540-30119-6_4.

[18] Michael Argyle and Mark Cook. *Gaze and Mutual Gaze*. English. Cambridge University Press Cambridge, Eng. ; New York, 1976, xi, 210 p. : ISBN: 0521208653.

[19] Reuben M. Aronson, Thiago Santini, Thomas C. Kübler, Enkelejda Kasneci, Siddhartha Srinivasa, and Henny Admoni. "Eye-Hand Behavior in Human-Robot Shared Manipulation". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. Chicago, IL, USA: Association for Computing Machinery, 2018, pp. 4–13. ISBN: 9781450349536. DOI: 10.1145/3171221.3171287.

[20] Ebrahim Babaei, Namrata Srivastava, Joshua Newn, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. "Faces of Focus: A Study on the Facial Cues of Attentional States". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376566.

[21] Thomas Bader and Jürgen Beyerer. "Natural Gaze Behavior as Input Modality for Human-Computer Interaction". In: *Eye Gaze in Intelligent User Interfaces: Gaze-based Analyses, Models and Applications*. Ed. by Yukiko I. Nakano, Cristina Conati, and Thomas Bader. London: Springer London, 2013, pp. 161–183. ISBN: 978-1-4471-4784-8. DOI: 10.1007/978-1-4471-4784-8_9.

[22] Gagan Bansal, Besmira Nushi, Ece Kamar, Walter Lasecki, Dan Weld, and Eric Horvitz. "Beyond Accuracy: The Role of Mental Models in Human-AI Team Performance". In: *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*. Vol. 7. 1. AAAI. Oct. 2019, pp. 2–11.

[23] Michael J. Barnes, Jessie Y. C. Chen, Kristin E. Schaefer, Troy Kelley, Cheryl Giammanco, and Susan Hill. "Five Requisites for Human-Agent Decision Sharing in Military Environments". In: *Advances in Human Factors in Robots and Unmanned Systems*. Cham: Springer International Publishing, 2017, pp. 39–48. ISBN: 978-3-319-41959-6. DOI: 10.1007/978-3-319-41959-6_4.

[24] Simon Baron-Cohen, Ruth Campbell, Annette Karmiloff-Smith, Julia Grant, and Jane Walker. "Are children with autism blind to the mentalistic significance of the eyes?" In: *British Journal of Developmental Psychology* 13.4 (1995), pp. 379–398. ISSN: 2044-835X. DOI: 10.1111/j.2044-835X.1995.tb00687.x.

[25] Simon Baron-Cohen, Alan M. Leslie, and Uta Frith. "Does the autistic child have a "theory of mind" ?" In: *Cognition* 21.1 (1985), pp. 37–46. ISSN: 0010-0277. DOI: https://doi.org/10.1016/0010-0277(85)90022-8.

[26] Roman Bednarik, Shahram Eivazi, and Hana Vrzakova. "A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data". In: *Eye Gaze in Intelligent User Interfaces: Gaze-based Analyses, Models and Applications*. Ed. by Yukiko I. Nakano, Cristina Conati, and Thomas Bader. London: Springer London, 2013, pp. 111–134. ISBN: 978-1-4471-4784-8. DOI: 10.1007/978-1-4471-4784-8_7.

[27] Kyle J. Behymer and John M. Flach. "From Autonomous Systems to Sociotechnical Systems: Designing Effective Collaborations". In: *She Ji: The Journal of Design, Economics, and Innovation* 2.2 (2016), pp. 105–114. ISSN: 2405-8726. DOI: https://doi.org/10.1016/j.sheji.2016.09.001.

[28] John Bowyer Bell and Barton Whaley. *Cheating and deception*. Transaction Publishers, 1991.

[29] Rachel K. E. Bellamy, Sean Andrist, Timothy Bickmore, Elizabeth F. Churchill, and Thomas Erickson. "Human-Agent Collaboration: Can an Agent Be a Partner?" In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 1289–1294. ISBN: 978-1-4503-4656-6. DOI: 10.1145/3027063.3051138.

[30] Or Biran and Courtenay Cotton. "Explanation and justification in machine learning: A survey". In: *IJCAI-17 Workshop on Explainable AI (XAI)*. Vol. 8. 1. 2017, pp. 8–13.

[31] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. "State-of-the-Art of Visualization for Eye Tracking Data". In: *EuroVis - STARs*. Ed. by R. Borgo, R. Maciejewski, and I. Viola. The Eurographics Association, 2014. ISBN: 978-3-03868-028-4. DOI: 10.2312/eurovisstar.20141173.

[32] Nate Blaylock and James Allen. "Statistical Goal Parameter Recognition". In: *Proceedings of the Fourteenth International Conference on International Conference on Automated Planning and Scheduling*. ICAPS'04. Whistler, British Columbia, Canada: AAAI Press, 2004, pp. 297–304. ISBN: 1-57735-200-9. URL: http://dl.acm.org/citation.cfm?id=3037008.3037047.

[33] Richard A. Bolt. ""Put-That-There": Voice and Gesture at the Graphics Interface". In: *SIGGRAPH Comput. Graph.* 14.3 (July 1980), pp. 262–270. ISSN: 0097-8930. DOI: 10.1145/965105.807503.

[34] Jean-David Boucher, Ugo Pattacini, Amelie Lelong, Gerard Bailly, Frederic Elisei, Sascha Fagel, Peter Dominey, and Jocelyne Ventre-Dominey. "I Reach Faster When I See You Look: Gaze Effects in Human–Human and Human–Robot Face-to-Face Cooperation". In: *Frontiers in Neurorobotics* 6 (2012), p. 3. ISSN: 1662-5218. DOI: 10.3389/fnbot.2012.00003.

[35] Myles Brand. "Intending and Acting: Toward a Naturalized Action Theory". In: *Journal of Philosophy* 84.1 (1984), pp. 49–54.

[36] Michael Bratman. *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information, 1987.

[37] Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. "Coordinating cognition: The costs and benefits of shared gaze during collaborative search". In: *Cognition* 106.3 (2008), pp. 1465–1477. ISSN: 0010-0277. DOI: http://dx.doi.org/10.1016/j.cognition.2007.05.012.

[38] Jeff Brewer, Sarah D'Angelo, and Darren Gergle. "Iris: Gaze Visualization Design Made Easy". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, D504:1–D504:4. ISBN: 978-1-4503-5621-3. DOI: 10.1145/3170427.3186502.

[39] Jeanne H. Brockmyer, Christine M. Fox, Kathleen A. Curtiss, Evan McBroom, Kimberly M. Burkhart, and Jacquelyn N. Pidruzny. "The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing". In: *Journal of Experimental Social Psychology* 45.4 (2009), pp. 624–634. ISSN: 0022-1031. DOI: 10.1016/j.jesp.2009.02.016.

[40] Noam Brown and Tuomas Sandholm. "Superhuman AI for heads-up no-limit poker: Libratus beats top professionals". In: *Science* 359.6374 (2018), pp. 418–424. ISSN: 0036-8075. DOI: 10.1126/science.aao1733.

[41] Frank Broz and Hagen Lehmann. "A Gaze Controller for Coordinating Mutual Gaze During Conversational Turn-Taking in Human-Robot Interaction". In: *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*. HRI'15 Extended Abstracts. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 131–132. ISBN: 978-1-4503-3318-4. DOI: 10.1145/2701973.2702043.

[42] Frank Broz, Hagen Lehmann, Yukiko Nakano, and Bilge Mutlu. "Gaze in HRI: From Modeling to Communication". In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '12. Boston, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 491–492. ISBN: 978-1-4503-1063-5. DOI: 10.1145/2157689.2157845.

[43] Ricardo Buettner. "Cognitive Workload of Humans Using Artificial Intelligence Systems: Towards Objective Measurement Applying Eye-Tracking Technology". In: *KI 2013: Advances in Artificial Intelligence*. Ed. by Ingo J. Timm and Matthias Thimm. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 37–48. ISBN: 978-3-642-40942-4. DOI: 10.1007/978-3-642-40942-4_4.

[44] Andreas Bulling, Daniel Roggen, and Gerhard Troester. "What's in the Eyes for Context-Awareness?" In: *IEEE Pervasive Computing* 10.2 (2011), pp. 48–57. DOI: 10.1109/MPRV.2010.49.

[45] Andreas Bulling, Jamie A. Ward, Hans Gellersen, and Gerhard. Troster. "Eye Movement Analysis for Activity Recognition Using Electrooculography". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.4 (Apr. 2011), pp. 741–753. ISSN: 0162-8828. DOI: 10.1109/TPAMI.2010.86.

[46] Mihai Burzo, Mohamed Abouelenien, Veronica Perez-Rosas, and Rada Mihalcea. "Multimodal Deception Detection". In: *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition - Volume 2*. Association for Computing Machinery and Morgan & Claypool, 2018, pp. 419–453. ISBN: 9781970001716. DOI: 10.1145/3107990.3108005.

[47] George Butterworth and Nicholas Jarrett. "What minds have in common is space: Spatial mechanisms serving joint visual attention in infancy". In: *British Journal of Developmental Psychology* 9.1 (1991), pp. 55–72. ISSN: 2044-835X. DOI: 10.1111/j.2044-835X.1991.tb00862.x.

[48] Lindsey Byom and Bilge Mutlu. "Theory of mind: mechanisms, methods, and new directions". In: *Frontiers in Human Neuroscience* 7 (2013), p. 413. ISSN: 1662-5161. DOI: 10.3389/fnhum.2013.00413.

[49] Andrew J. Calder, Andrew D. Lawrence, Jill Keane, Sophie K. Scott, Adrian M. Owen, Ingrid Christoffels, and Andrew W. Young. "Reading the mind from eye gaze". In: *Neuropsychologia* 40.8 (2002), pp. 1129–1138. ISSN: 0028-3932. DOI: 10.1016/S0028-3932(02)00008-8.

[50] Murray Campbell, A.Joseph Hoane, and Feng-hsiung Hsu. "Deep Blue". In: *Artificial Intelligence* 134.1 (2002), pp. 57–83. ISSN: 0004-3702. DOI: https://doi.org/10.1016/S0004-3702(01)00129-1.

[51] Sandra Carberry. "Techniques for Plan Recognition". In: *User Modeling and User-Adapted Interaction* 11.1 (Mar. 2001), pp. 31–48. ISSN: 1573-1391. DOI: 10.1023/A:1011118925938.

[52] John M. Carroll. "Five reasons for scenario-based design". In: *Interacting with Computers* 13.1 (2000), pp. 43–60. ISSN: 0953-5438. DOI: https://doi.org/10.1016/S0953-5438(00)00023-0.

[53] John M. Carroll, ed. *Scenario-Based Design: Envisioning Work and Technology in System Development*. USA: John Wiley & Sons, Inc., 1995. ISBN: 0471076597.

[54] Marcus Carter, Joshua Newn, Eduardo Velloso, and Frank Vetere. "Remote Gaze and Gesture Tracking on the Microsoft Kinect: Investigating the Role of Feedback". In: *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*. OzCHI '15. Parkville, VIC, Australia: Association for Computing Machinery, 2015, pp. 167–176. ISBN: 978-1-4503-3673-4. DOI: 10.1145/2838739.2838778.

[55] Jessie Y. C. Chen and Michael J. Barnes. "Human–Agent Teaming for Multirobot Control: A Review of Human Factors Issues". In: *IEEE Transactions on Human-Machine Systems* 44.1 (Feb. 2014), pp. 13–29. ISSN: 2168-2305. DOI: 10.1109/THMS.2013.2293535.

[56] Jessie Y. C. Chen and Michael J. Barnes. "Supervisory Control of Multiple Robots: Effects of Imperfect Automation and Individual Differences". In: *Human Factors* 54.2 (2012). PMID: 22624284, pp. 157–174. DOI: 10.1177/0018720811435843.

[57] Jessie Y. C. Chen and Michael J. Barnes. "Supervisory control of multiple robots in dynamic tasking environments". In: *Ergonomics* 55.9 (2012). PMID: 22676776, pp. 1043–1058. DOI: 10.1080/00140139.2012.689013.

[58] Jessie Y. C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Wright, and Michael Barnes. "Situation awareness-based agent transparency and human-autonomy teaming effectiveness". In: *Theoretical Issues in Ergonomics Science* 19.3 (2018), pp. 259–282. DOI: 10.1080/1463922X.2017.1315750.

[59] Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. "Deixis and Gaze in Collaborative Work at a Distance (over a Shared Map): A Computational Model to Detect Misunderstandings". In: *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications*. ETRA '08. Savannah, Georgia: Association for Computing Machinery, 2008, pp. 173–180. ISBN: 9781595939821. DOI: 10.1145/1344471.1344515.

[60] Leigh Clark, Philip Doyle, Diego Garaialde, Emer Gilmartin, Stephan Schlögl, Jens Edlund, Matthew Aylett, João Cabral, Cosmin Munteanu, Justin Edwards, and Benjamin R. Cowan. "The State of Speech in HCI: Trends, Themes and Challenges". In: *Interacting with Computers* 31.4 (Sept. 2019), pp. 349–371. ISSN: 0953-5438. DOI: 10.1093/iwc/iwz016.

[61] Philip R. Cohen, C. Raymond Perrault, and James F. Allen. *Beyond question answering*. Vol. 245274. ed Hillsdale, NJ: Lawrence Erlbaum Associates, 1981.

[62] Mark Cook. "Gaze and Mutual Gaze in Social Encounters: How long—and when—we look others "in the eye" is one of the main signals in nonverbal communication". In: *American Scientist* 65.3 (1977), pp. 328–333. ISSN: 00030996. URL: http://www.jstor.org/stable/27847843.

[63] Sarah D'Angelo and Andrew Begel. "Improving Communication Between Pair Programmers Using Shared Gaze Awareness". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. CHI '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 6245–6290. ISBN: 978-1-4503-4655-9. DOI: 10.1145/3025453.3025573.

[64] Sarah D'Angelo, Jeff Brewer, and Darren Gergle. "Iris: A Tool for Designing Contextually Relevant Gaze Visualizations". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA '19. Denver, Colorado: Association for Computing Machinery, 2019, 79:1–79:5. ISBN: 978-1-4503-6709-7. DOI: 10.1145/3317958.3318228.

[65] Sarah D'Angelo and Darren Gergle. "An Eye For Design: Gaze Visualizations for Remote Collaborative Work". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, 349:1–349:12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173923.

[66] Sarah D'Angelo and Darren Gergle. "Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. Santa Clara, California, USA: Association for Computing Machinery, 2016, pp. 2492–2496. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858499.

[67] Sergey Demyanov, James Bailey, Kotagiri Ramamohanarao, and Christopher Leckie. "Detection of Deception in the Mafia Party Game". In: *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ICMI '15. Seattle, Washington, USA: Association for Computing Machinery, 2015, pp. 335–342. ISBN: 978-1-4503-3912-4. DOI: 10.1145/2818346.2820745.

[68] Virgina Dignum. "The Role of Organization in Agent Systems". In: *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*. IGI Global, 2009, pp. 1–16. DOI: 10.4018/978-1-60566-256-5.ch001.

[69] Jonathan Dodge, Sean Penney, Claudia Hilderbrand, Andrew Anderson, and Margaret Burnett. "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*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, 562:1–562:12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3174136.

[70] Angela E. Douglas. *Symbiotic Interactions*. Oxford Science Publications. Oxford University Press, 1994. ISBN: 0198542941.

[71] Andrew T. Duchowski. *Eye Tracking Methodology: Theory and Practice*. 3rd. Springer Publishing Company, Incorporated, 2017. ISBN: 3319578812.

[72] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. "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*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, 282:1–282:13. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173856.

[73] Andrew T. Duchowski, Margaux M. Price, Miriah Meyer, and Pilar Orero. "Aggregate Gaze Visualization with Real-time Heatmaps". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: Association for Computing Machinery, 2012, pp. 13–20. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168558.

[74] James F. Dunnigan and Albert A. Nofi. *Victory and Deceit: Deception and Trickery at War*. Writers Club Press, 2001. ISBN: 9780595184057.

[75] Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. "Bringing Transparency Design into Practice". In: *23rd International Conference on Intelligent User Interfaces*. IUI '18. Tokyo, Japan: Association for Computing Machinery, 2018, pp. 211–223. ISBN: 978-1-4503-4945-1. DOI: 10.1145/3172944.3172961.

[76] Shiri Einav and Bruce M Hood. "Tell-tale eyes: children's attribution of gaze aversion as a lying cue." In: *Developmental Psychology* 44.6 (2008), p. 1655. DOI: 10.1037/a0013299.

[77] Paul Ekman and Wallace V. Friesen. "Nonverbal Leakage and Clues to Deception". In: *Psychiatry* 32.1 (1969). PMID: 27785970, pp. 88–106. DOI: 10.1080/00332747.1969.11023575.

[78] Mica R. Endsley. "Toward a Theory of Situation Awareness in Dynamic Systems". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37.1 (1995), pp. 32–64. ISSN: 00187208. DOI: 10.1518/001872095779049543.

[79] Xiaocong Fan, Michael McNeese, Bingjun Sun, Timothy Hanratty, Laurel Allender, and John Yen. "Human–Agent Collaboration for Time-Stressed Multicontext Decision Making". In: *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 40.2 (Mar. 2010), pp. 306–320. ISSN: 1558-2426. DOI: 10.1109/TSMCA.2009.2035302.

[80] Stephen Fiore, Travis Wiltshire, Emilio Lobato, Florian Jentsch, Wesley Huang, and Benjamin Axelrod. "Toward understanding social cues and signals in human–robot interaction: effects of robot gaze and proxemic behavior". In: *Frontiers in Psychology* 4 (2013), p. 859. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2013.00859. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2013.00859.

[81] Paul M. Fitts. *Human Engineering for an Effective Air-navigation and Traffic-control System*. National Research Council, Div. of, 1951.

[82] Ross Flom, Kang Lee, and Darwin Muir. *Gaze Following: Its Development and Significance*. Mahwah, NJ : Lawrence Erlbaum Associates, 2007., Dec. 2007. ISBN: 0805847502.

[83] Tom Foulsham and Maria Lock. "How the Eyes Tell Lies: Social Gaze During a Preference Task". In: *Cognitive Science* 39.7 (2015), pp. 1704–1726. ISSN: 1551-6709. DOI: 10.1111/cogs.12211.

[84] Alejo Freire, Michelle Eskritt, and Kang Lee. "Are Eyes Windows to a Deceiver's Soul? Children's Use of Another's Eye Gaze Cues in a Deceptive Situation." In: *Developmental Psychology* 40.6 (2004), pp. 1093–1104. ISSN: 0012-1649. DOI: 10.1037/0012-1649.40.6.1093.

[85] Alexandra Frischen, Andrew P. Bayliss, and Steven P. Tipper. "Gaze Cueing of Attention: Visual Attention, Social Cognition, and Individual Differences". In: *Psychological Bulletin* 133.4 (2007), p. 694. DOI: 10.1037/0033-2909.133.4.694.

[86] Limin Paul Fu, James Landay, Michael Nebeling, Yingqing Xu, and Chen Zhao. "Redefining Natural User Interface". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, pp. 1–3. ISBN: 9781450356213. DOI: 10.1145/3170427.3190649.

[87] Helen L. Gallagher and Christopher D. Frith. "Functional imaging of 'theory of mind'". In: *Trends in Cognitive Sciences* 7.2 (2003), pp. 77–83. ISSN: 1364-6613. DOI: 10.1016/S1364-6613(02)00025-6.

[88] Hector Geffner and Blai Bonet. "A Concise Introduction to Models and Methods for Automated Planning". In: *Synthesis Lectures on Artificial Intelligence and Machine Learning* 7.2 (2013), pp. 1–141. DOI: 10.2200/S00513ED1V01Y201306AIM022.

[89] Michail N. Giannakos, Kshitij Sharma, Ilias O. Pappas, Vassilis Kostakos, and Eduardo Velloso. "Multimodal data as a means to understand the learning experience". In: *International Journal of Information Management* 48 (2019), pp. 108–119. ISSN: 0268-4012. DOI: 10.1016/j.ijinfomgt.2019.02.003.

[90] Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. "GeoGazemarks: Providing Gaze History for the Orientation on Small Display Maps". In: *Proceedings of the 14th ACM International Conference on Multimodal Interaction*. ICMI '12. Santa Monica, California, USA: Association for Computing Machinery, 2012, pp. 165–172. ISBN: 978-1-4503-1467-1. DOI: 10.1145/2388676.2388711.

[91] Matthias S. Gobel, Heejung S. Kim, and Daniel C. Richardson. "The dual function of social gaze". In: *Cognition* 136.Supplement C (2015), pp. 359–364. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2014.11.040.

[92] Joseph H. Goldberg and Jonathan I. Helfman. "Visual Scanpath Representation". In: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*. ETRA '10. Austin, Texas: Association for Computing Machinery, 2010, pp. 203–210. ISBN: 978-1-60558-994-7. DOI: 10.1145/1743666.1743717.

[93] Susan Goldin-Meadow. "The Role of Gesture in Communication and Thinking". In: *Trends in Cognitive Sciences* 3.11 (1999), pp. 419–429. DOI: 10.1016/s1364-6613(99)01397-2.

[94] Miriam Greis, Jessica Hullman, Michael Correll, Matthew Kay, and Orit Shaer. "Designing for Uncertainty in HCI: When Does Uncertainty Help?" In: *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '17. Denver, Colorado, USA: Association for Computing Machinery, 2017, pp. 593–600. ISBN: 9781450346566. DOI: 10.1145/3027063.3027091.

[95] Barbara J. Grosz. "Collaborative Systems (AAAI-94 Presidential Address)". In: *AI Magazine* 17.2 (Mar. 1996), p. 67. DOI: 10.1609/aimag.v17i2.1223. URL: https://www.aaai.org/ojs/index.php/aimagazine/article/view/1223.

[96] Jonathan Grudin. "From Tool to Partner: The Evolution of Human-Computer Interaction". In: *Synthesis Lectures on Human-Centered Informatics* 10.1 (2017), pp. i–183. DOI: 10.2200/S00745ED1V01Y201612HCI035.

[97] Nader Hanna and Deborah Richards. "A Collaborative Agent Architecture with Human-Agent Communication Model". In: *Cognitive Agents for Virtual Environments*. Ed. by Frank Dignum, Cyril Brom, Koen Hindriks, Martin Beer, and Deborah Richards. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 70–88. ISBN: 978-3-642-36444-0. DOI: 10.1007/978-3-642-36444-0_5.

[98] Maaike Harbers, Karel van den Bosch, and John-Jules Ch. Meyer. "A Study into Preferred Explanations of Virtual Agent Behavior". In: *Intelligent Virtual Agents*. Ed. by Zsofia Ruttkay, Michael Kipp, Anton Nijholt, and Hannes Hogni Vilhjalmsson. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 132–145. ISBN: 978-3-642-04380-2. DOI: 10.1007/978-3-642-04380-2_17.

[99] Sandra G. Hart and Lowell E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In: *Human Mental Workload*. Ed. by Peter A. Hancock and Najmedin Meshkati. Vol. 52. Advances in Psychology. North-Holland, 1988, pp. 139–183. DOI: 10.1016/S0166-4115(08)62386-9.

[100] Bradley Hayes and Julie A. Shah. "Improving Robot Controller Transparency Through Autonomous Policy Explanation". In: *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. Vienna, Austria: Association for Computing Machinery, 2017, pp. 303–312. ISBN: 978-1-4503-4336-7. DOI: 10.1145/2909824.3020233.

[101] Jane Heal. "Joint Attention and Understanding the Mind". In: *Oxford University PressJoint Attention: Communication and Other Minds*. Ed. by N. Elian, Christoph Hoerl, Teresa McCormack, and Johannes Roessler. Oxford University Press, 2005, pp. 34–44.

[102] Clinton Heinze. "Modelling Intention Recognition for Intelligent Agent Systems". PhD Thesis, The University of Melbourne. 2003. URL: www.dsto.defence.gov.au/corporate/reports/DSTO-RR-0286.pdf.

[103] Tove Helldin, Ulrika Ohlander, Göran Falkman, and Maria Riveiro. "Transparency of Automated Combat Classification". In: *Engineering Psychology and Cognitive Ergonomics*. Ed. by Don Harris. Cham: Springer International Publishing, 2014, pp. 22–33. ISBN: 978-3-319-07515-0. DOI: 10.1007/978-3-319-07515-0_3.

[104] Keita Higuch, Ryo Yonetani, and Yoichi Sato. "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*. CHI '16. San Jose, California, USA: Association for Computing Machinery, 2016, pp. 5180–5190. ISBN: 978-1-4503-3362-7. DOI: 10.1145/2858036.2858438.

[105] Sébastien Hillaire, Anatole Lécuyer, Gaspard Breton, and Tony Regia Corte. "Gaze Behavior and Visual Attention Model when Turning in Virtual Environments". In: *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. VRST '09. Kyoto, Japan: Association for Computing Machinery, 2009, pp. 43–50. ISBN: 978-1-60558-869-8. DOI: 10.1145/1643928.1643941.

[106] Robert R. Hoffman, Shane T. Mueller, Gary Klein, and Jordan Litman. "Metrics for Explainable AI: Challenges and Prospects". In: *arXiv preprint arXiv:1812.04608* (2018).

[107] Eric Horvitz and Tim Paek. "A Computational Architecture for Conversation". In: *Proceedings of the Seventh International Conference on User Modeling*. UM '99. Banff, Canada: Springer-Verlag New York, Inc., 1999, pp. 201–210. ISBN: 3-211-83151-7. URL: http://dl.acm.org/citation.cfm?id=317328.317354.

[108] Chien-Ming Huang, Sean Andrist, Allison Sauppé, and Bilge Mutlu. "Using gaze patterns to predict task intent in collaboration". In: *Frontiers in Psychology* 6 (2015), p. 1049. ISSN: 1664-1078. DOI: 10.3389/fpsyg.2015.01049.

[109] Chien-Ming Huang and Bilge Mutlu. "Anticipatory Robot Control for Efficient Human-Robot Collaboration". In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. HRI '16. Christchurch, New Zealand: IEEE Press, 2016, pp. 83–90. ISBN: 9781467383707. DOI: 10.1109/HRI.2016.7451737.

[110] Hsiao-Ying Huang and Masooda Bashir. "Personal Influences on Dynamic Trust Formation in Human-Agent Interaction". In: *Proceedings of the 5th International Conference on Human Agent Interaction*. HAI '17. Bielefeld, Germany: Association for Computing Machinery, 2017, pp. 233–243. ISBN: 9781450351133. DOI: 10.1145/3125739.3125749.

[111] Xuedong Huang, James Baker, and Raj Reddy. "A Historical Perspective of Speech Recognition". In: *Commun. ACM* 57.1 (Jan. 2014), pp. 94–103. ISSN: 0001-0782. DOI: 10.1145/2500887.

[112] Carina Huchler. "An MCTS agent for Ticket to Ride". Master Thesis, Maastricht University, 2015. URL: https://project.dke.maastrichtuniversity.nl/games/files/msc/Huchler_thesis.pdf.

[113] Robert Hyman and Daniel Druckman. "A broader concept of deception". In: *In the mind's eye: Enhancing Human Performance* (1991).

[114] Ryo Ishii, Ryota Ooko, Yukiko I. Nakano, and Tokoaki Nishida. "Effectiveness of Gaze-Based Engagement Estimation in Conversational Agents". In: *Eye Gaze in Intelligent User Interfaces: Gaze-based Analyses, Models and Applications*. Ed. by Yukiko I. Nakano, Cristina Conati, and Thomas Bader. London: Springer London, 2013, pp. 85–110. ISBN: 978-1-4471-4784-8. DOI: 10.1007/978-1-4471-4784-8_6.

[115] Laurent Itti, Christof Koch, and Ernst Niebur. "A model of saliency-based visual attention for rapid scene analysis". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.11 (1998), pp. 1254–1259.

[116] Robert J. K. Jacob. "The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get". In: *ACM Trans. Inf. Syst.* 9.2 (Apr. 1991), pp. 152–169. ISSN: 1046-8188. DOI: 10.1145/123078.128728.

[117] Robert J. K. Jacob. "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*. CHI '90. Seattle, Washington, USA: Association for Computing Machinery, 1990, pp. 11–18. ISBN: 0-201-50932-6. DOI: 10.1145/97243.97246.

[118] Dinesh Babu Jayagopi and Jean-Marc Odobez. "Given That, Should i Respond? Contextual Addressee Estimation in Multi-Party Human-Robot Interactions". In: *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*. HRI '13. Tokyo, Japan: IEEE Press, 2013, pp. 147–148. ISBN: 9781467330558. DOI: 10.1109/HRI.2013.6483544.

[119] Patrick Jermann and Marc-Antoine Nüssli. "Effects of Sharing Text Selections on Gaze Cross-recurrence and Interaction Quality in a Pair Programming Task". In: *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*. CSCW '12. Seattle, Washington, USA: Association for Computing Machinery, 2012, pp. 1125–1134. ISBN: 978-1-4503-1086-4. DOI: 10.1145/2145204.2145371.

[120] Philip N. Johnson-Laird. *Mental models: Towards a cognitive science of language, inference, and consciousness*. 6. Harvard University Press, 1983.

[121] Catholijn M. Jonker, M. Birna van Riemsdijk, and Bas Vermeulen. "Shared Mental Models". In: *Coordination, Organizations, Institutions, and Norms in Agent Systems VI*. Ed. by Marina De Vos, Nicoletta Fornara, Jeremy V. Pitt, and George Vouros. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 132–151. ISBN: 978-3-642-21268-0. DOI: 10.1007/978-3-642-21268-0_8.

[122] Marcel Adam Just and Patricia A. Carpenter. "A theory of Reading: From Eye Fixations to Comprehension". In: *Psychological Review* 87.4 (1980). DOI: 10.1037/0033-295X.87.4.329.

[123] Froduald Kabanza, Philipe Bellefeuille, Francis Bisson, Abder Rezak Benaskeur, and Hengameh Irandoust. "Opponent Behaviour Recognition for Real-time Strategy Games". In: *Proceedings of the 5th AAAI Conference on Plan, Activity, and Intent Recognition*. AAAIWS'10-05. AAAI Press, 2010, pp. 29–36. URL: http://dl.acm.org/citation.cfm?id=2908558.2908563.

[124] Daniel Kahneman and Patrick Egan. *Thinking, fast and slow*. Vol. 1. Farrar, Straus and Giroux New York, 2011.

[125] Clare-Marie Karat, Christine Halverson, Daniel Horn, and John Karat. "Patterns of Entry and Correction in Large Vocabulary Continuous Speech Recognition Systems". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '99. Pittsburgh, Pennsylvania, USA: Association for Computing Machinery, 1999, pp. 568–575. ISBN: 0201485591. DOI: 10.1145/302979.303160.

[126] John Karat. "Scenario Use in the Design of a Speech Recognition System". In: *Scenario-Based Design: Envisioning Work and Technology in System Development*. USA: John Wiley & Sons, Inc., 1995, pp. 109–133. ISBN: 0471076597.

[127] John Karat, Jennifer Lai, Catalina Danis, and Catherine Wolf. "Speech User Interface Evolution". In: *Human Factors and Voice Interactive Systems*. Ed. by Daryle Gardner-Bonneau. Boston, MA: Springer US, 1999, pp. 1–35. ISBN: 978-1-4757-2980-1. DOI: 10.1007/978-1-4757-2980-1_1.

[128] Daphne E. Karreman, Gilberto U. Sepúlveda Bradford, Betsy E.M.A.G. van Dijk, Manja Lohse, and Vanessa Evers. "What Happens When a Robot Favors Someone? How a Tour Guide Robot Uses Gaze Behavior to Address Multiple Persons While Storytelling about Art". In: *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*. HRI '13. Tokyo, Japan: IEEE Press, 2013, pp. 157–158. ISBN: 9781467330558.

[129] Robert Kass and Tim Finin. *The Need for User Models in Generating Expert System Explanations*. Tech. rep. University of Pennsylvania, June 1988, pp. 1–32. URL: http://repository.upenn.edu/cis%5C_reports/585.

[130] Henry A. Kautz and James F. Allen. "Generalized Plan Recognition". In: *Proceedings of the Fifth AAAI National Conference on Artificial Intelligence*. AAAI'86. Philadelphia, Pennsylvania: AAAI Press, 1986, pp. 32–37. URL: http://dl.acm.org/citation.cfm?id=2887770.2887776.

[131] Frank C. Keil. "Explanation and Understanding". In: *Annual Review of Psychology* 57.1 (2006), pp. 227–254. DOI: 10.1146/annurev.psych.57.102904.190100.

[132] J. F. Kelley. "An Iterative Design Methodology for User-friendly Natural Language Office Information Applications". In: *ACM Trans. Inf. Syst.* 2.1 (Jan. 1984), pp. 26–41. ISSN: 1046-8188. DOI: 10.1145/357417.357420.

[133] Adam Kendon. "Some functions of gaze-direction in social interaction". In: *Acta Psychologica* 26 (1967), pp. 22–63. ISSN: 0001-6918. DOI: 10.1016/0001-6918(67)90005-4.

[134] Peter Kiefer. *Challenges in Gaze-based Intention Recognition*. en. Dagstuhl, Germany, 2018. DOI: 10.3929/ethz-b-000311201.

[135] Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. "Using Eye Movements to Recognize Activities on Cartographic Maps". In: *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. SIGSPATIAL'13. Orlando, Florida: Association for Computing Machinery, 2013, pp. 488–491. ISBN: 978-1-4503-2521-9. DOI: 10.1145/2525314.2525467.

[136] Nathan Kirchner, Alen Alempijevic, and Gamini Dissanayake. "Nonverbal Robot-Group Interaction Using an Imitated Gaze Cue". In: *Proceedings of the 6th International Conference on Human-Robot Interaction*. HRI '11. Lausanne, Switzerland: Association for Computing Machinery, 2011, pp. 497–504. ISBN: 9781450305617. DOI: 10.1145/1957656.1957824. URL: https://doi.org/10.1145/1957656.1957824.

[137] Gary Klien, David D. Woods, Jeffery M. Bradshaw, Robert R. Hoffman, and Paul J. Feltovich. "Ten challenges for making automation a "team player" in joint human-agent activity". In: *IEEE Intelligent Systems* 19.6 (Nov. 2004), pp. 91–95. ISSN: 1541-1672. DOI: 10.1109/MIS.2004.74.

[138] Mark L Knapp, Judith A Hall, and Terrence G Horgan. *Nonverbal Communication in Human Interaction*. Cengage Learning, 2013.

[139] Hiromi Kobayashi and Shiro Kohshima. "Evolution of the Human Eye as a Device for Communication". In: *Primate Origins of Human Cognition and Behavior*. Ed. by Tetsuro Matsuzawa. Tokyo: Springer Japan, 2001, pp. 383–401. ISBN: 978-4-431-09423-4. DOI: 10.1007/978-4-431-09423-4_19.

[140] Hiromi Kobayashi and Shiro Kohshima. "Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye". In: *Journal of Human Evolution* 40.5 (2001), pp. 419–435. ISSN: 0047-2484. DOI: 10.1006/jhev.2001.0468.

[141] Daphne Koller and Avi Pfeffer. "Generating and Solving Imperfect Information Games". In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI'95. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995, pp. 1185–1192. ISBN: 1-55860-363-8. URL: http://dl.acm.org/citation.cfm?id=1643031.1643053.

[142] Robert E. Kraut, Susan R. Fussell, and Jane Siegel. "Visual Information As a Conversational Resource in Collaborative Physical Tasks". In: *Hum.-Comput. Interact.* 18.1 (June 2003), pp. 13–49. ISSN: 0737-0024. DOI: 10.1207/S15327051HCI1812_2. URL: https://doi-org/10.1207/S15327051HCI1812_2.

[143] Grete Helena Kütt, Kevin Lee, Ethan Hardacre, and Alexandra Papoutsaki. "Eye-Write: Gaze Sharing for Collaborative Writing". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, 497:1–497:12. ISBN: 978-1-4503-5970-2. DOI: 10.1145/3290605.3300727.

[144] Michael F. Land. "Vision, eye movements, and natural behavior". In: *Visual Neuroscience* 26.1 (2009), pp. 51–62. DOI: 10.1017/S0952523808080899.

[145] Michael F. Land and Benjamin Tatler. *Looking and acting: vision and eye movements in natural behaviour*. Oxford University Press, 2009. DOI: 10.1093/acprof:oso/9780198570943.001.0001.

[146] Pat Langley, Ben Meadows, Mohan Sridharan, and Dongkyu Choi. "Explainable Agency for Intelligent Autonomous Systems". In: *Innovative Applications of Artificial Intelligence*. 2017. URL: https://www.aaai.org/ocs/index.php/IAAI/IAAI17/paper/view/15046.

[147] Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. "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*. ACE2016. Osaka, Japan: Association for Computing Machinery, 2016, 17:1–17:9. ISBN: 978-1-4503-4773-0. DOI: 10.1145/3001773.3001774.

[148] Michael Lankes, Joshua Newn, Bernhard Maurer, Eduardo Velloso, Martin Dechant, and Hans Gellersen. "EyePlay Revisited: Past, Present and Future Challenges for Eye-Based Interaction in Games". In: *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. CHI PLAY '18 Extended Abstracts. Melbourne, VIC, Australia: Association for Computing Machinery, 2018, pp. 689–693. ISBN: 9781450359689. DOI: 10.1145/3270316.3271549.

[149] Neal Lesh, Joe Marks, Charles Rich, and Candace L. Sidner. "'Man-Computer Symbiosis' Revisited: Achieving Natural Communication and Collaboration with Computers". In: *IEICE Transactions on Electronics* (Dec. 2004). URL: http://www.merl.com/publications/TR2004-053.

[150] Neal Lesh, Charles Rich, and Candace L. Sidner. "Using Plan Recognition in Human-computer Collaboration". In: *Proceedings of the Seventh International Conference on User Modeling*. UM '99. Banff, Canada: Springer-Verlag New York, Inc., 1999, pp. 23–32. ISBN: 3-211-83151-7. URL: http://dl.acm.org/citation.cfm?id=317328.317331.

[151] Alan M. Leslie. "Pretense and representation: The origins of "theory of mind."" In: *Psych. review* 94.4 (1987), p. 412. DOI: 10.1037/0033-295X.94.4.412.

[152] Timothy R. Levine. "Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection". In: *Journal of Language and Social Psychology* 33.4 (2014), pp. 378–392. DOI: 10.1177/0261927X14535916.

[153] Jerry Li, Mia Manavalan, Sarah D'Angelo, and Darren Gergle. "Designing Shared Gaze Awareness for Remote Collaboration". In: *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion*. CSCW '16 Companion. San Francisco, California, USA: Association for Computing Machinery, 2016, pp. 325–328. ISBN: 978-1-4503-3950-6. DOI: 10.1145/2818052.2869097.

[154] Songpo Li and Xiaoli Zhang. "Implicit Intention Communication in Human–Robot Interaction Through Visual Behavior Studies". In: *IEEE Transactions on Human-Machine Systems* 47.4 (Aug. 2017), pp. 437–448. ISSN: 2168-2291. DOI: 10.1109/THMS.2017.2647882.

[155] Joseph Carl Robnett Licklider. "Man-Computer Symbiosis". In: *IRE Transactions on Human Factors in Electronics* HFE-1.1 (Mar. 1960), pp. 4–11. ISSN: 0099-4561. DOI: 10.1109/THFE2.1960.4503259.

[156] Brian Y. Lim and Anind K. Dey. "Investigating Intelligibility for Uncertain Context-aware Applications". In: *Proceedings of the 13th International Conference on Ubiquitous Computing*. UbiComp '11. Beijing, China: Association for Computing Machinery, 2011, pp. 415–424. ISBN: 978-1-4503-0630-0. DOI: 10.1145/2030112.2030168.

[157] Darren Lim. "Taking Students out for a Ride: Using a Board Game to Teach Graph Theory". In: *SIGCSE Bull.* 39.1 (Mar. 2007), pp. 367–371. ISSN: 0097-8418. DOI: 10.1145/1227504.1227439.

[158] David V. Lu and William D. Smart. "Towards more efficient navigation for robots and humans". In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Nov. 2013, pp. 1707–1713. DOI: 10.1109/IROS.2013.6696579.

[159] Prashan Madumal. "Explainable Agency in Intelligent Agents: Doctoral Consortium". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. Montreal QC, Canada: International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 2432–2434. ISBN: 978-1-4503-6309-9. URL: http://dl.acm.org/citation.cfm?id=3306127.3332137.

[160] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. "A Grounded Interaction Protocol for Explainable Artificial Intelligence". In: *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '19. Montreal QC, Canada: International Foundation for Autonomous Agents and Multiagent Systems, 2019, pp. 1033–1041. ISBN: 978-1-4503-6309-9. URL: http://dl.acm.org/citation.cfm?id=3306127.3331801.

[161] Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. "Explainable Reinforcement Learning through a Causal Lens". In: *The Thirty-Fourth AAAI Conference on Artificial Intelligence*. AAAI 2020. AAAI Press, 2020, pp. 2493–2500. DOI: 10.1609/aaai.v34i03.5631.

[162] Prashan Madumal, Ronal Singh, Joshua Newn, and Frank Vetere. "Interaction Design for Explainable AI: Workshop Proposal". In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. OzCHI '18. In: *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. OzCHI '18. Melbourne, Australia: Association for Computing Machinery, 2018, pp. 607–608. ISBN: 978-1-4503-6188-0. DOI: 10.1145/3292147.3293450.

[163] Paivi Majaranta, Hirotaka Aoki, Mick Donegan, Dan Witzner Hansen, and John Paulin Hansen. *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*. 1st. Hershey, PA: Information Science Reference - Imprint of: IGI Publishing, 2011. ISBN: 161350098X, 9781613500989.

[164] Bertram F. Malle and Joshua Knobe. "Which behaviors do people explain? A basic actor–observer asymmetry." In: *Journal of Personality and Social Psychology* 72.2 (1997), p. 288. DOI: 10.1037/0022-3514.72.2.288.

[165] Samantha Mann, Aldert Vrij, Sharon Leal, Pär Anders Granhag, Lara Warmelink, and Dave Forrester. "Windows to the Soul? Deliberate Eye Contact as a Cue to Deceit". In: *Journal of Nonverbal Behavior* 36.3 (2012), pp. 205–215. ISSN: 1573-3653. DOI: 10.1007/s10919-012-0132-y.

[166] Peta Masters and Sebastian Sardina. "Cost-Based Goal Recognition for Path-Planning". In: *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*. AAMAS '17. Sao Paulo, Brazil: International Foundation for Autonomous Agents and Multiagent Systems, 2017, pp. 750–758. URL: http://dl.acm.org/citation.cfm?id=3091125.3091232.

[167] Oludamilare Matthews, Zhanna Sarsenbayeva, Weiwei Jiang, Joshua Newn, Eduardo Velloso, Sarah Clinch, and Jorge Goncalves. "Inferring the Mood of a Community From Their Walking Speed: A Preliminary Study". In: *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*. UbiComp '18. Singapore, Singapore: Association for Computing Machinery, 2018, pp. 1144–1149. ISBN: 9781450359665. DOI: 10.1145/3267305.3274759.

[168] Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. "EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game". In: *Entertainment Computing - ICEC 2016: 15th IFIP TC 14 International Conference, Vienna, Austria, September 28-30, 2016, Proceedings*. Ed. by Günter Wallner, Simone Kriglstein, Helmut Hlavacs, Rainer Malaka, Artur Lugmayr, and Hyun-Seung Yang. Cham: Springer International Publishing, 2016, pp. 102–114. ISBN: 978-3-319-46100-7. DOI: 10.1007/978-3-319-46100-7_9.

[169] Nikolaos Mavridis. "A review of verbal and non-verbal human–robot interactive communication". In: *Robotics and Autonomous Systems* 63 (2015), pp. 22–35. ISSN: 0921-8890. DOI: 10.1016/j.robot.2014.09.031.

[170] Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth André. "Exploring a Model of Gaze for Grounding in Multimodal HRI". In: *Proceedings of the 16th International Conference on Multimodal Interaction*. ICMI '14. Istanbul, Turkey: Association for Computing Machinery, 2014, pp. 247–254. ISBN: 9781450328852. DOI: 10.1145/2663204.2663275.

[171] Albert Mehrabian. *Nonverbal Communication*. Routledge, 2017.

[172] Alfred R. Mele. *Springs of Action: Understanding Intentional Behavior*. Oxford University Press, 1992.

[173] Fernando de Mesentier Silva, Scott Lee, Julian Togelius, and Andy Nealen. "AI-based Playtesting of Contemporary Board Games". In: *Proceedings of the 12th International Conference on the Foundations of Digital Games*. FDG '17. Hyannis, Massachusetts: Association for Computing Machinery, 2017, 13:1–13:10. ISBN: 978-1-4503-5319-9. DOI: 10.1145/3102071.3102105.

[174] Tim Miller. "Explanation in Artificial Intelligence: Insights from the Social Sciences". In: *Artificial Intelligence* 267 (2019), pp. 1–38. ISSN: 0004-3702. DOI: 10.1016/j.artint.2018.07.007.

[175] Andrew F. Monk and Caroline Gale. "A Look Is Worth a Thousand Words: Full Gaze Awareness in Video-Mediated Conversation". In: *Discourse Processes* 33.3 (2002), pp. 257–278. DOI: 10.1207/S15326950DP3303_4.

[176] AJung Moon, Daniel M. Troniak, Brian Gleeson, Matthew K.X.J. Pan, Minhua Zheng, Benjamin A. Blumer, Karon MacLean, and Elizabeth A. Croft. "Meet Me Where I'm Gazing: How Shared Attention Gaze Affects Human-robot Handover Timing". In: *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*. HRI '14. Bielefeld, Germany: Association for Computing Machinery, 2014, pp. 334–341. ISBN: 978-1-4503-2658-2. DOI: 10.1145/2559636.2559656.

[177] Chris Moore and Philip J. Dunham. *Joint Attention: Its Origins and Role in Development*. Lawrence Erlbaum, 1995.

[178] Johanna D. Moore and Cécile L. Paris. "Requirements for an expert system explanation facility". In: *Computational Intelligence* 7.4 (1991), pp. 367–370. DOI: 10.1111/j.1467-8640.1991.tb00409.x.

[179] Christian Muise, Vaishak Belle, Paolo Felli, Sheila McIlraith, Tim Miller, Adrian R. Pearce, and Liz Sonenberg. "Planning over Multi-Agent Epistemic States: A Classical Planning Approach". In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI'15. Austin, Texas: AAAI Press, 2015, pp. 3327–3334. ISBN: 0262511290.

[180] Jorge Munoz, Georgios N. Yannakakis, Fiona Mulvey, Dan W. Hansen, German Gutierrez, and Araceli Sanchis. "Towards gaze-controlled platform games". In: *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*. Aug. 2011, pp. 47–54. DOI: 10.1109/CIG.2011.6031988.

[181] Bilge Mutlu, Fumitaka Yamaoka, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Nonverbal Leakage in Robots: Communication of Intentions Through Seemingly Unintentional Behavior". In: *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. HRI '09. La Jolla, California, USA: Association for Computing Machinery, 2009, pp. 69–76. ISBN: 978-1-60558-404-1. DOI: 10.1145/1514095.1514110.

[182] Vignesh Narayanan, Yu Zhang, Nathaniel Mendoza, and Subbarao Kambhampati. "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*. HRI'15 Extended Abstracts. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 161–162. ISBN: 978-1-4503-3318-4. DOI: 10.1145/2701973.2702042.

[183] Joshua Newn. "Enabling Intent Recognition Through Gaze Awareness in User Interfaces". In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI EA '18. Montreal QC, Canada: Association for Computing Machinery, 2018, DC15:1–DC15:4. DOI: 10.1145/3170427.3173028.

[184] Joshua Newn, Fraser Allison, Eduardo Velloso, and Frank Vetere. "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*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018, 261:1–261:12. ISBN: 978-1-4503-5620-6. DOI: 10.1145/3173574.3173835.

[185] Joshua Newn, Ronal Singh, Fraser Allison, Prashan Madumal, Eduardo Velloso, and Frank Vetere. "Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents". In: *Human-Computer Interaction – INTERACT 2019*. Ed. by David Lamas, Fernando Loizides, Lennart Nacke, Helen Petrie, Marco Winckler, and Panayiotis Zaphiris. Cham: Springer International Publishing, 2019, pp. 255–281. ISBN: 978-3-030-29384-0. DOI: 10.1007/978-3-030-29384-0_17.

[186] Joshua Newn, Ronal Singh, Eduardo Velloso, and Frank Vetere. "Combining Implicit Gaze and AI for Real-Time Intention Projection". In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. UbiComp/ISWC '19 Adjunct. London, United Kingdom: Association for Computing Machinery, 2019, pp. 324–327. ISBN: 9781450368698. DOI: 10.1145/3341162.3343786.

[187] Joshua Newn, Benjamin Tag, Ronal Singh, Eduardo Velloso, and Frank Vetere. "AI-Mediated Gaze-Based Intention Recognition for Smart Eyewear: Opportunities & Challenges". In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM*

International Symposium on Wearable Computers. UbiComp/ISWC '19 Adjunct. London, United Kingdom: Association for Computing Machinery, 2019, pp. 637–642. ISBN: 9781450368698. DOI: 10.1145/3341162.3348387.

[188] Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. "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*. CHI PLAY '17. Amsterdam, The Netherlands: Association for Computing Machinery, 2017, pp. 541–552. ISBN: 978-1-4503-4898-0. DOI: 10.1145/3116595.3116624.

[189] Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. "Exploring the Effects of Gaze Awareness on Multiplayer Gameplay". In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. CHI PLAY Companion '16. Austin, Texas, USA: Association for Computing Machinery, 2016, pp. 239–244. DOI: 10.1145/2968120.2987740.

[190] Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. "Multimodal Segmentation on a Large Interactive Tabletop: Extending Interaction on Horizontal Surfaces with Gaze". In: *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*. ISS '16. Niagara Falls, Ontario, Canada: Association for Computing Machinery, 2016, pp. 251–260. ISBN: 9781450342483. DOI: 10.1145/2992154.2992179.

[191] Raymond S. Nickerson. "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises". In: *Review of General Psychology* 2.2 (1998), pp. 175–220. DOI: 10.1037/1089-2680.2.2.175.

[192] Ingrid Nunes and Dietmar Jannach. "A Systematic Review and Taxonomy of Explanations in Decision Support and Recommender Systems". In: *User Modeling and User-Adapted Interaction* 27.3-5 (Dec. 2017), pp. 393–444. ISSN: 0924-1868. DOI: 10.1007/s11257-017-9195-0.

[193] Raquel Oliveira, Patricia Arriaga, Patricia Alves-Oliveira, Filipa Correia, Sofia Petisca, and Ana Paiva. "Friends or Foes? Socioemotional Support and Gaze Behaviors in Mixed Groups of Humans and Robots". In: *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '18. Chicago, IL, USA: Association for Computing Machinery, 2018, pp. 279–288. ISBN: 9781450349536. DOI: 10.1145/3171221.3171272.

[194] Santiago Ontañón, Gabriel Synnaeve, Alberto Uriarte, Florian Richoux, David Churchill, and Mike Preuss. "A Survey of Real-Time Strategy Game AI Research and Competition in StarCraft". In: *IEEE Transactions on Computational Intelligence and AI in Games* 5.4 (Dec. 2013), pp. 293–311. DOI: 10.1109/TCIAIG.2013.2286295.

[195] Elisabeth Pacherie. "The phenomenology of action: A conceptual framework". In: *Cognition* 107.1 (2008), pp. 179–217. ISSN: 0010-0277. DOI: 10.1016/j.cognition.2007.09.003.

[196] Ramon Pereira, Nir Oren, and Felipe Meneguzzi. "Landmark-Based Heuristics for Goal Recognition". In: *AAAI*. 2017. URL: https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14666.

[197] Ken Pfeuffer, Jason Alexander, and Hans Gellersen. "GazeArchers: Playing with Individual and Shared Attention in a Two-player Look&Shoot Tabletop Game". In: *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia*. MUM '16. Rovaniemi, Finland: Association for Computing Machinery, 2016, pp. 213–216. ISBN: 978-1-4503-4860-7. DOI: 10.1145/3012709.3012717.

[198] Klaus Pohl and Peter Haumer. "Modelling Contextual Information about Scenarios". In: *Proceedings of the Third International Workshop on Requirements Engineering: Foundations of Software Quality REFSQ'97*. 1997, pp. 187–204.

[199] Martin Porcheron, Joel E. Fischer, Stuart Reeves, and Sarah Sharples. "Voice Interfaces in Everyday Life". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. CHI '18. Montreal QC, Canada: Association for Computing Machinery, 2018. ISBN: 9781450356206. DOI: 10.1145/3173574.3174214.

[200] David Premack and G. Woodruff. "Does the Chimpanzee Have a Theory of Mind?" In: *Behavioral and Brain Sciences* 4.4 (1978), pp. 515–629. DOI: 10.1017/S0140525X00076512.

[201] Pernilla Qvarfordt, David Beymer, and Shumin Zhai. "RealTourist: A Study of Augmenting Human-human and Human-computer Dialogue with Eye-gaze Overlay". In: *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*. INTERACT'05. Rome, Italy: Springer-Verlag, 2005, pp. 767–780. ISBN: 3-540-28943-7, 978-3-540-28943-2. DOI: 10.1007/11555261_61.

[202] Pernilla Qvarfordt and Shumin Zhai. "Conversing with the User Based on Eye-gaze Patterns". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '05. Portland, Oregon, USA: Association for Computing Machinery, 2005, pp. 221–230. ISBN: 1-58113-998-5. DOI: 10.1145/1054972.1055004.

[203] Sarvapali D. Ramchurn, Joel E. Fischer, Yuki Ikuno, Feng Wu, Jack Flann, and Antony Waldock. "A Study of Human-Agent Collaboration for Multi-UAV". In: *Proceedings of the 24th International Conference on Artificial Intelligence*. IJCAI'15. Buenos Aires, Argentina: AAAI Press, 2015, pp. 1184–1192. ISBN: 9781577357384.

[204] Miquel Ramirez and Hector Geffner. "Probabilistic Plan Recognition Using Off-the-shelf Classical Planners". In: *Proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence*. AAAI'10. Atlanta, Georgia: AAAI Press, 2010, pp. 1121–1126. URL: http://dl.acm.org/citation.cfm?id=2898607.2898786.

[205] Vincent van Rheden, Bernhard Maurer, Dorothé Smit, Martin Murer, and Manfred Tscheligi. "LaserViz: Shared Gaze in the Co-Located Physical World". In: *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*. TEI '17. Yokohama, Japan: Association for Computing Machinery, 2017, pp. 191–196. ISBN: 978-1-4503-4676-4. DOI: 10.1145/3024969.3025010.

[206] Charles Rich, Brett Ponsler, Aaron Holroyd, and Candance L. Sidner. "Recognizing engagement in human-robot interaction". In: *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Mar. 2010, pp. 375–382. DOI: 10.1109/HRI.2010.5453163.

[207] Laurel D. Riek. "Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines". In: *J. Hum.-Robot Interact.* 1.1 (July 2012), pp. 119–136. ISSN: 2163-0364. DOI: 10.5898/JHRI.1.1.Riek.

[208] Verena Rieser. *Bootstrapping Reinforcement Learning-Based Dialogue Strategies from Wizard-of-Oz Data*. English. Vol. XXVIII. This thesis has been awarded with the Eduard-Martin prize for distinguished doctoral dissertations. German Research Center for Artificial Intelligence, 2008.

[209] Cornelis Joost van Rijsbergen. *Information Retrieval*. 2nd. Newton, MA, USA: Butterworth-Heinemann, 1979. ISBN: 0408709294.

[210] Melissa J. Rogerson, Martin R. Gibbs, and Wally Smith. "What Can We Learn from Eye Tracking Boardgame Play?" In: *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '17 Extended Abstracts. Amsterdam, The Netherlands: Association for Computing Machinery, 2017, pp. 519–526. ISBN: 978-1-4503-5111-9. DOI: 10.1145/3130859.3131314.

[211] William B. Rouse and Nancy M. Morris. "On looking into the black box: Prospects and limits in the search for mental models." In: *Psychological Bulletin* 100.3 (1986), p. 349. DOI: 10.1037/0033-2909.100.3.349.

[212] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd. USA: Prentice Hall Press, 2009. ISBN: 0136042597.

[213] Katie Salen and Eric Zimmerman. *Rules of Play: Game Design Fundamentals*. The MIT Press, 2003. ISBN: 0262240459, 9780262240451.

[214] Akanksha Saran, Elaine Schaertl Short, Andrea Thomaz, and Scott Niekum. "Understanding Teacher Gaze Patterns for Robot Learning". In: vol. 100. Proceedings of Machine Learning Research. PMLR, 2020, pp. 1247–1258. URL: http://proceedings.mlr.press/v100/saran20a.html.

[215] Brian Scassellati. "Theory of Mind for a Humanoid Robot". In: *Autonomous Robots* 12.1 (Jan. 2002), pp. 13–24. ISSN: 1573-7527. DOI: 10.1023/A:1013298507114.

[216] Kristin E. Schaefer, Edward R. Straub, Jessie Y. C. Chen, Joe Putney, and A. W. Evans. "Communicating intent to develop shared situation awareness and engender trust in human-agent teams". In: *Cognitive Systems Research* 46 (2017). Situation Awareness in Human-Machine Interactive Systems, pp. 26–39. ISSN: 1389-0417. DOI: 10.1016/j.cogsys.2017.02.002.

[217] Albrecht Schmidt. *Implicit Human Computer Interaction Through Context*. Tech. rep. 2. Personal Technologies, 2000, pp. 191–199. DOI: https://doi.org/10.1007/BF01324126.

[218] Albrecht Schmidt and Thomas Herrmann. "Intervention User Interfaces: A New Interaction Paradigm for Automated Systems". In: *Interactions* 24.5 (Aug. 2017), pp. 40–45. ISSN: 1072-5520. DOI: 10.1145/3121357.

[219] John R. Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, 2012.

[220] Marwin H. S. Segler, Mike Preuss, and Mark P. Waller. "Planning chemical syntheses with deep neural networks and symbolic AI". In: *Nature* 555.7698 (2018), pp. 604–610. DOI: 10.1038/nature25978.

[221] Stephen Shepherd. "Following Gaze: Gaze-Following Behavior as a Window into Social Cognition". In: *Frontiers in Integrative Neuroscience* 4 (2010), p. 5. ISSN: 1662-5145. DOI: 10.3389/fnint.2010.00005.

[222] Ivan Shindev, Yu Sun, Michael Coovert, Jenny Pavlova, and Tiffany Lee. "Exploration of Intention Expression for Robots". In: *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction*. HRI '12. Boston, Massachusetts, USA: Association for Computing Machinery, 2012, pp. 247–248. ISBN: 978-1-4503-1063-5. DOI: 10.1145/2157689.2157778.

[223] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. "Mastering the Game of Go without Human Knowledge". In: *Nature* 550.7676 (2017), pp. 354–359. DOI: 10.1038/nature24270.

[224] Ronal Singh, Tim Miller, Joshua Newn, Liz Sonenberg, Eduardo Velloso, and Frank Vetere. "Combining Planning with Gaze for Online Human Intention Recognition". In: *Proceedings of the 17th International Conference on Autonomous Agents and Multiagent System*. AAMAS '18. Stockholm, Sweden: International Foundation for Autonomous Agents and Multiagent Systems, 2018. URL: http://dl.acm.org/citation.cfm?id=3237383.3237457.

[225] Ronal Singh, Tim Miller, Joshua Newn, Eduardo Velloso, Frank Vetere, and Liz Sonenberg. "Combining gaze and AI planning for online human intention recognition". In: *Artificial Intelligence* 284 (2020), p. 103275. ISSN: 0004-3702. DOI: 10.1016/j.artint.2020.103275.

[226] Oleg Špakov. "Comparison of Eye Movement Filters Used in HCI". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: Association for Computing Machinery, 2012, pp. 281–284. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168616.

[227] Oleg Špakov, Howell Istance, Kari-Jouko Räihä, Tiia Viitanen, and Harri Siirtola. "Eye Gaze and Head Gaze in Collaborative Games". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA '19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3317959.3321489.

[228] Namrata Srivastava, Sadia Nawaz, Jason M. Lodge, Eduardo Velloso, Sarah Erfani, and James Bailey. "Exploring the Usage of Thermal Imaging for Understanding Video Lecture Designs and Students' Experiences". In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. LAK '20. Frankfurt, Germany: Association for Computing Machinery, 2020, pp. 250–259. ISBN: 9781450377126. DOI: 10.1145/3375462.3375514.

[229] Namrata Srivastava, Joshua Newn, and Eduardo Velloso. "Combining Low and Mid-Level Gaze Features for Desktop Activity Recognition". In: *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2.4 (Dec. 2018). DOI: 10.1145/3287067.

[230] Namrata Srivastava, Eduardo Velloso, Jason M. Lodge, Sarah Erfani, and James Bailey. "Continuous Evaluation of Video Lectures from Real-Time Difficulty Self-Report". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019, pp. 1–12. ISBN: 9781450359702. DOI: 10.1145/3290605.3300816.

[231] Randy Stein and Susan E. Brennan. "Another Person's Eye Gaze As a Cue in Solving Programming Problems". In: *Proceedings of the 6th International Conference on Multimodal Interfaces*. ICMI '04. State College, PA, USA: Association for Computing Machinery, 2004, pp. 9–15. ISBN: 1-58113-995-0. DOI: 10.1145/1027933.1027936.

[232] Kimberly Stowers, Nicholas Kasdaglis, Michael Rupp, Jessie Y. C. Chen, Daniel Barber, and Michael J. Barnes. "Insights into Human-Agent Teaming: Intelligent Agent Transparency and Uncertainty". In: *Advances in Human Factors in Robots and Unmanned Systems*. Cham: Springer International Publishing, 2017, pp. 149–160. ISBN: 978-3-319-41959-6. DOI: 10.1007/978-3-319-41959-6_13.

[233] Gita Sukthankar, Christopher Geib, Hung Hai Bui, David Pynadath, and Robert P. Goldman. *Plan, Activity, and Intent Recognition: Theory and Practice*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2014. ISBN: 0123985323, 9780123985323.

[234] Karim A. Tahboub. "Intelligent Human-Machine Interaction Based on Dynamic Bayesian Networks Probabilistic Intention Recognition". In: *J. Intell. Robotics Syst.* 45.1 (Jan. 2006), pp. 31–52. ISSN: 0921-0296. DOI: 10.1007/s10846-005-9018-0.

[235] Milind Tambe and Paul S. Rosenbloom. "RESC: An Approach for Real-time, Dynamic Agent Tracking". In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 1*. IJCAI'95. Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc., 1995, pp. 103–110. ISBN: 1-55860-363-8, 978-1-558-60363-9. URL: http://dl.acm.org/citation.cfm?id=1625855.1625869.

[236] David Tennenhouse. "Proactive Computing". In: *Commun. ACM* 43.5 (May 2000), pp. 43–50. ISSN: 0001-0782. DOI: 10.1145/332833.332837. URL: https://doi.org/10.1145/332833.332837.

[237] Loren G. Terveen. "Overview of Human-Computer Collaboration". In: *Knowledge-Based Systems* 8.2 (1995). Human-computer collaboration, pp. 67–81. ISSN: 0950-7051. DOI: 10.1016/0950-7051(95)98369-H.

[238] Serge Thill and Tom Ziemke. "The Role of Intentions in Human-Robot Interaction". In: *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '17. Vienna, Austria: Association for Computing Machinery, 2017, pp. 427–428. ISBN: 978-1-4503-4885-0. DOI: 10.1145/3029798.3029802.

[239] Michael Tomasello, Brian Hare, Hagen Lehmann, and Josep Call. "Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis". In: *Journal of Human Evolution* 52.3 (2007), pp. 314–320. ISSN: 0047-2484. DOI: 10.1016/j.jhevol.2006.10.001.

[240] Jayson Turner, Eduardo Velloso, Hans Gellersen, and Veronica Sundstedt. "EyePlay: Applications for Gaze in Games". In: *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play*. CHI PLAY '14. Toronto, Ontario, Canada: Association for Computing Machinery, 2014, pp. 465–468. ISBN: 978-1-4503-3014-5. DOI: 10.1145/2658537.2659016.

[241] Vaibhav V. Unhelkar, Przemslaw A. Lasota, Quirin Tyroller, Rares-Darius Buhai, Laurie Marceau, Barbara Deml, and Julie A. Shah. "Human-Aware Robotic Assistant for Collaborative Assembly: Integrating Human Motion Prediction With Planning in Time". In: *IEEE Robotics and Automation Letters* 3.3 (July 2018), pp. 2394–2401. ISSN: 2377-3766. DOI: 10.1109/LRA.2018.2812906.

[242] Dominique Vaufreydaz, Wafa Johal, and Claudine Combe. "Starting engagement detection towards a companion robot using multimodal features". In: *Robotics and Autonomous Systems* 75 (2016). Assistance and Service Robotics in a Human Environment, pp. 4–16. ISSN: 0921-8890. DOI: https://doi.org/10.1016/j.robot.2015.01.004.

[243] Eduardo Velloso and Marcus Carter. "The Emergence of EyePlay: A Survey of Eye Interaction in Games". In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. CHI PLAY '16. Austin, Texas, USA: Association for Computing Machinery, 2016, pp. 171–185. ISBN: 978-1-4503-4456-2. DOI: 10.1145/2967934.2968084.

[244] Mor Vered, Pierce Howe, Tim Miller, Liz Sonenberg, and Eduardo Velloso. "Demand-Driven Transparency for Monitoring Intelligent Agents". In: *IEEE Transactions on Human-Machine Systems* 50.3 (2020), pp. 264–275. DOI: 10.1109/THMS.2020.2988859.

[245] Mor Vered, Gal A. Kaminka, and Sivan Biham. "Online Goal Recognition through Mirroring: Humans and Agents". In: *Proceedings of the Fourth Annual Conference on Advances in Cognitive Systems*. A slightly modified version appears in Proceedings of the IJCAI 2016 workshop on Human-Agent Interaction Design and Models (HAIDM). 2016.

[246] Frank Vetere, Kenton O'Hara, Jeni Paay, Bernd Ploderer, Richard Harper, and Abigail Sellen. "Social NUI: Social Perspectives in Natural User Interfaces". In: *Proceedings of the 2014 Companion Publication on Designing Interactive Systems*. DIS Companion '14. Vancouver, BC, Canada: Association for Computing Machinery, 2014, pp. 215–218. ISBN: 9781450329033. DOI: 10.1145/2598784.2598802.

[247] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. "Designing Theory-Driven User-Centric Explainable AI". In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. CHI '19. Glasgow, Scotland Uk: Association for Computing Machinery, 2019. ISBN: 9781450359702. DOI: 10.1145/3290605.3300831.

[248] Haofei Wang and Bertram E. Shi. "Gaze Awareness Improves Collaboration Efficiency in a Collaborative Assembly Task". In: *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ETRA '19. Denver, Colorado: Association for Computing Machinery, 2019. ISBN: 9781450367097. DOI: 10.1145/3317959.3321492.

[249] Ning Wang, David V Pynadath, and Susan G Hill. "Trust calibration within a human-robot team: Comparing automatically generated explanations". In: *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*. IEEE Press. 2016, pp. 109–116. DOI: 10.1109/HRI.2016.7451741.

[250] Nadir Weibel, Adam Fouse, Colleen Emmenegger, Sara Kimmich, and Edwin Hutchins. "Let's Look at the Cockpit: Exploring Mobile Eye-tracking for Observational Research on the Flight Deck". In: *Proceedings of the Symposium on Eye Tracking Research and Applications*. ETRA '12. Santa Barbara, California: Association for Computing Machinery, 2012, pp. 107–114. ISBN: 978-1-4503-1221-9. DOI: 10.1145/2168556.2168573.

[251] Mark Weiser. "The Computer for the 21st Century". In: *Human-Computer Interaction: Toward the Year 2000*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1995, pp. 933–940. ISBN: 1558602461.

[252] Douglas Brent West. *Introduction to Graph Theory*. Prentice Hall, 2001. ISBN: 9780130144003.

[253] Stefanie Wetzel, Katharina Spiel, and Sven Bertel. "Dynamically Adapting an AI Game Engine Based on Players' Eye Movements and Strategies". In: *Proceedings of the 2014 ACM SIGCHI Symposium on Engineering Interactive Computing Systems*. EICS '14. Rome, Italy: Association for Computing Machinery, 2014, pp. 3–12. ISBN: 978-1-4503-2725-1. DOI: 10.1145/2607023.2607029.

[254] Barton Whaley. "Toward a General Theory of Deception". In: *Journal of Strategic Studies* 5.1 (1982), pp. 178–192. DOI: 10.1080/01402398208437106.

[255] Daniel Wigdor and Dennis Wixon. *Brave NUI World: Designing Natural User Interfaces for Touch and Gesture*. 1st. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2011. ISBN: 0123822319.

[256] Alan F. T. Winfield. "Experiments in Artificial Theory of Mind: From Safety to Story-Telling". In: *Frontiers in Robotics and AI* 5 (2018), p. 75. ISSN: 2296-9144. DOI: 10.3389/frobt.2018.00075.

[257] Christine T. Wolf. "Explainability Scenarios: Towards Scenario-Based XAI Design". In: *Proceedings of the 24th International Conference on Intelligent User Interfaces*. IUI '19. Marina del Ray, California: Association for Computing Machinery, 2019, pp. 252–257. ISBN: 9781450362726. DOI: 10.1145/3301275.3302317.

[258] Niels Wouters, Ryan Kelly, Eduardo Velloso, Katrin Wolf, Hasan Shahid Ferdous, Joshua Newn, Zaher Joukhadar, and Frank Vetere. "Biometric Mirror: Exploring Ethical Opinions towards Facial Analysis and Automated Decision-Making". In: *Proceedings of the 2019 on Designing Interactive Systems Conference*. DIS '19. San Diego, CA, USA: Association for Computing Machinery, 2019, pp. 447–461. ISBN: 9781450358507. DOI: 10.1145/3322276.3322304.

[259] Tian (Linger) Xu, Hui Zhang, and Chen Yu. "See You See Me: The Role of Eye Contact in Multimodal Human-Robot Interaction". In: *ACM Trans. Interact. Intell. Syst.* 6.1 (May 2016), 2:1–2:22. ISSN: 2160-6455. DOI: 10.1145/2882970.

[260] Qian Yang, Aaron Steinfeld, Carolyn Rosé, and John Zimmerman. "Re-Examining Whether, Why, and How Human-AI Interaction Is Uniquely Difficult to Design". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. CHI '20. Honolulu, HI, USA: Association for Computing Machinery, 2020, pp. 1–13. ISBN: 9781450367080. DOI: 10.1145/3313831.3376301.

[261] Georgios N. Yannakakis and Julian Togelius. "Introduction". In: *Artificial Intelligence and Games*. Cham: Springer International Publishing, 2018, pp. 3–28. ISBN: 978-3-319-63519-4. DOI: 10.1007/978-3-319-63519-4_1.

[262] Shumin Zhai. "What's in the Eyes for Attentive Input". In: *Commun. ACM* 46.3 (Mar. 2003), pp. 34–39. ISSN: 0001-0782. DOI: 10.1145/636772.636795.

[263] Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. "Look Together: Using Gaze for Assisting Co-located Collaborative Search". In: *Personal Ubiquitous Comput.* 21.1 (Feb. 2017), pp. 173–186. ISSN: 1617-4909. DOI: 10.1007/s00779-016-0969-x.

[264] Qiushi Zhou, Joshua Newn, Namrata Srivastava, Jorge Goncalves, Tilman Dingler, and Eduardo Velloso. "Cognitive Aid: Task Assistance Based On Mental Workload Estimation". In: *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. CHI '19 Extended Abstracts. Glasgow, Scotland, UK: Association for Computing Machinery, 2019. DOI: 10.1145/3290607.3313010.

[265] Qiushi Zhou, Joshua Newn, Benjamin Tag, Hao-Ping Lee, Chaofan Wang, and Eduardo Velloso. "Ubiquitous Smart Eyewear Interactions Using Implicit Sensing and Unobtrusive Information Output". In: *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*. UbiComp/ISWC '19 Adjunct. London, United Kingdom: Association for Computing Machinery, 2019, pp. 661–666. ISBN: 9781450368698. DOI: 10.1145/3341162.3348392.

A

Published Papers

A.1 Paper: *Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games*

DOI: 10.1145/3116595.3116624

Evaluating Real-Time Gaze Representations to Infer Intentions in Competitive Turn-Based Strategy Games

Joshua Newn[1], **Eduardo Velloso**[1], **Fraser Allison**[1], **Yomna Abdelrahman**[2], **Frank Vetere**[1]
[1]Microsoft Research Centre for Social NUI, University of Melbourne, Australia
[2]Institute for Visualization and Interactive Systems, University of Stuttgart, Germany
[1]{newnj,evelloso,fallison,f.vetere}@unimelb.edu.au, [2]yomna.abdelrahman@vis.uni-stuttgart.de

ABSTRACT
In this paper, we investigate nine different visual representations of gaze in a competitive digital game setting. We evaluate the ability of spectators to infer a player's intentions in the game for each visual representation. Our results show that spectators have a remarkable ability to infer intent accurately using all nine visualizations, but that visualizations with certain characteristics were more comprehensible and more readily revealed the player's intent. The real-time *Heatmap* visualization was the most highly preferred by participants and the most effective in revealing intent, due to its ability to balance real-time gaze information with a persistent summary of recent gaze behaviour. Our findings show that eye-tracking visualization can enable playful interactions in competitive games based on players' ability to interpret opponents' attention and intention through gaze information.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces: Input devices and strategies

Author Keywords
Eye tracking; Gaze; Nonverbal leakage; Shared gaze; Gaze awareness; Competitive Gaming; Intent Prediction

INTRODUCTION
Players' nonverbal signals are an important element in many competitive board and card games. By monitoring an opponent's body language, facial expressions and gaze direction, a player can infer their next moves or call their bluffs. Gaze is a strong signal of a person's area of attention and interest [9], and can give away unintended information—what psychology researchers call 'nonverbal leakage' [8]. Gaze is equally important in collaborative settings, as a cue for collaborators to monitor each other's attention and coordinate their actions [14]. Despite its importance in co-located settings, shared gaze awareness is typically absent in remote multiplayer games. In this paper, we demonstrate an opportunity to bring shared gaze awareness into this setting by comparing nine different ways of visualising gaze within a digital multiplayer version of a popular board game—*Ticket to Ride*[1].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI PLAY '17, October 15-18, 2017, Amsterdam, Netherlands
© 2017 ACM. ISBN 978-1-4503-4898-0/17/10…$15.00
DOI: https://doi.org/10.1145/3116595.3116624

Figure 1: An opponent's gaze presented as a real-time Heatmap over *Ticket to Ride game* showing interest in the east coast of the North America. This is an animated figure and it is best viewed in Adobe Reader.

The advantages in observing a partner's gaze in a *collaborative* task are clear, leading researchers to build multiple systems where users' gaze points are streamed over a network to improve coordination [3]. However, the opportunities for gaze in *competitive* settings are more nuanced. Recent studies have explored these possibilities by using visualizations of players' gaze to open up novel gameplay experiences and increase social presence [16, 20]. Normally in remote multiplayer games, a player's only clues for predicting their opponents' current and future plans are the in-game actions they have previously taken; players do not have access to the bodily cues (such as gaze) that enrich the experience of games played around a table or in a shared space. When given information about an opponent's gaze in a networked game, a player can build hypotheses about what their opponent is thinking through observation of the areas or game elements that receive the most visual attention. This creates further opportunities for a player to deliberately manipulate their own gaze behaviour to deceive an opponent about their plans and goals. This contrast between the tendency for observed gaze to reveal intent and the ability of a player to use their eye movements deceptively makes shared gaze awareness a rich resource for game design.

[1]https://www.daysofwonder.com/online/en/t2r/

However, studies on shared gaze visualizations for various applications have found that the introduction of shared gaze awareness can be *disruptive*, often being referred to as 'distracting' and 'confusing' [6, 16]. We hypothesise that this is caused by two primary problems with gaze visualization for digital interfaces. First, the way in which gaze has been visually represented in past works may be inadequate or ill-suited for the task and setting. This is expected, as the design of visual representations in graphical user interfaces is highly dependent on their context of use. This problem is exacerbated by an inherent problem with gaze: gaze is 'always on', which makes designing gaze representations difficult as the technology cannot distinguish between focused scrutiny and more absent-minded looking.

Second, humans are not accustomed to interpreting visual representations of gaze, as the focal point of gaze is 'invisible' in normal interpersonal interaction. Gaze information can be 'noisy' due to the jittery movements and the way the eyes work (fixation-saccade-fixation cycle), and this is reflected in the representations. Gaze data is difficult to interpret and often misunderstood [6], as even trained professionals in eye-tracking evaluations make use of a range of gaze visualizations to make sense of the same collected data. Further, as gaze is fast-moving and never entirely still, any real-time visualization overlaid on a user interface adds a highly distracting element that may interfere with the viewer's own thought process.

To effectively use gaze awareness in a game, the designer must understand how the choice of gaze visualization will affect players' understanding and enjoyment. Inspired by how players observe the gaze of one another in tabletop boardgames, we selected a popular board game—*Ticket to Ride*—for the purposes of our research. Using recorded clips from eye-tracked gameplay, we assessed nine gaze representations in three different stages of the game (early, mid, late). This paper serves as an initial novel groundwork for determining the characteristics of gaze representations that allow human players to make inferences about other players' intentions and into derivation of potential insights into their strategies (goals and plans).

RELATED WORK

In this section, we summarise related works that require a user to interpret gaze representations of another, which falls into two key areas: research on *shared gaze*, and works that compare and interpret the use of gaze representations. We have chosen to exclude works related to eye-tracking evaluations in which gaze data is interpreted by trained professionals, as our interest is specifically on the ability of non-specialist users to interpret visual gaze representations.

Shared Gaze

Shared gaze systems are those that give users the ability to see the gaze of a partner, typically to improve communication and coordination between pairs in collaborative problem-solving tasks. Here, the eyes are typically not directly visible, and the addition of a gaze visualization provides a complementary layer of nonverbal communication. Knowing where someone is looking can provide rich information, from inference about their intentions to clues about their current cognitive activity (based on behaviours such as scanning, focused interest in an object, and repeated comparison of different objects) [24]. With this knowledge, much research in HCI (Human-Computer Interaction) has used gaze in collaborative settings to foster coordination. For instance, Stein and Brennan [24] found an improvement in problem-solving task performance when another person's gaze was used as a cue. Gaze has been shown to be effective in co-located settings, e.g. when multiple users perform a visual search task on a large display [29]. A common approach is to present a user's gaze as a cursor overlaid on the shared visual display space, to create an awareness of that user's focus area [5, 24].

Gaze has shown promise when used implicitly with other modalities in remote communication. Qvarfordt and Zhai applied gaze in a dialogue system, where a remote assistant is able to detect the interest of a remote user through their gaze patterns while conversing with them (remote user is unaware that their gaze is being tracked) [22]. Tracking gaze behaviour can serve as a natural deictic pointer and thus reveal interest in a natural way. Brennan et al. have demonstrated that shared gaze can be more efficient than speech for the rapid communication of spatial information (spatial referencing) [3].

Despite the numerous benefits mentioned, gaze as an input suffers from inherent problems is that gaze is 'always on', related to the 'Midas Touch' problem that leads to accidental triggering of gaze-enabled interface widgets [12]. To combat this, gaze is often paired with another modality, such as voice interaction; however, this can also lead to confusion when there is a mismatch between gaze and voice inputs [6]. However, in competitive settings, this ambiguity can work as a resource for game design, revealing a glimpse into players' intentions without giving it away entirely.

Gaze Representation

How gaze information is represented strongly influences gaze interaction [4]. We reviewed two recent works in HCI that compare and contrast gaze representations to gather design implications for our study. Zhang et al. evaluate gaze representations for use in co-located collaboration, where two users share a screen and two gaze points are represented [29]. From a review of prior works, the authors selected four different gaze visualization formats for testing—*Cursor*, *Trajectory*, *Highlight* and *Spotlight*—of which participants preferred highlight and spotlight. The authors determined that attaining a balance between *visibility* and *distraction* was one of the biggest challenges when designing and selecting gaze visualizations.

In another work, Li et al. designed and evaluated two gaze representations aimed at improving the coordination between pairs of users [17]: *zoom focus* and *gaze trail*. Both representations used fixations processed from the gaze stream; the former aimed to show the precise location of a fixation, while the latter aimed to show the most recent gaze locations. *Zoom focus* was found to be cognitively demanding and *gaze trail* was found to have excessive information. The authors conclude that when designing gaze representation, one should consider *distraction* and *visual information*.

Games Context and Summary

Gaze is an emerging input in computer games, with a focus on using a player's own gaze as a direct control mechanism to en-

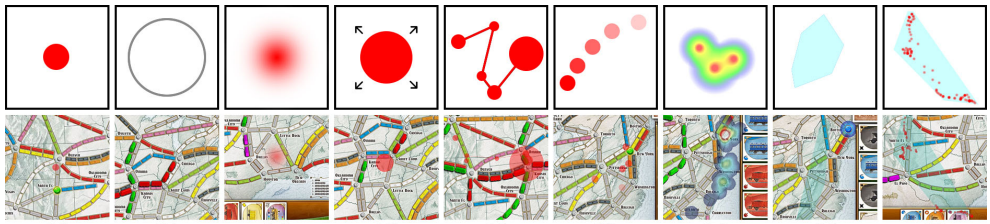


Figure 2: Top: The nine real-time gaze representations selected. Bottom: How each representation appears over the game *Ticket to Ride*. L-R: (a) Dot, (b) Cursor, (c) Spotlight, (d) Fixation, (e) Scanpath, (f) Fixation Trail, (g) Heatmap, (h) Convex Hull and (i) Bee Swarm.

able novel playful experiences [25, 26]. There is growing interest in implementing gaze as a mechanic for multiplayer games, in both cooperative [19, 21] and competitive [16, 20] settings. While the use of gaze in cooperative gameplay can directly leverage the findings of 'shared gaze' literature (e.g. [21]), competitive gameplay presents a different design challenge, as players have opposing objectives and are more likely to see different information on screen. We seek to find a gaze representation for use in competitive games that maintains the game balance, giving neither player a disproportionate edge over their opponent while not being disruptive to gameplay.

The existing literature makes clear that there is a need to balance *visibility*, *distraction* and *visual information*. Furthermore, we note that the problem with a majority of gaze representations is that they are designed to be used in post-processing of gaze data. Therefore, there is a need to design and evaluate novel gaze representations, such as that presented by Li et al. [17], to cater for settings where gaze input is not commonly used. Before we can design novel gaze representations, we need to understand the intrinsic characteristics of existing representations in relation to the context of use. In the next section, we present a variety of visual gaze representations that we have selected for use in our study.

REAL-TIME GAZE REPRESENTATIONS

Figure 2 shows the nine visual gaze representations which were evaluated in this study. For each representation, we provide an account of its current applications, its characteristics and the parameters for our use in the study. Most of the representations are common in existing in eye-tracking software, such as Tobii Pro Studio[2]. These representations are primarily used for gaze data analysis and to replay collected gaze data. We also considered a wide selection of representations from works that: (1) survey gaze representations (e.g. [2, 18]), (2) evaluate gaze representations (e.g. [29]) and (3) use gaze for interaction (e.g. [21]). In making our selection, we required the representations to be clearly distinguishable from one another. The selection was based on the characteristics *conspicuousness* and *visual information* (see Table 1). We chose the term *conspicuousness* (rather than 'visibility') to highlight the degree to which the representations can be used to display and draw attention, beyond just how well the viewer is able to see the representation. *Visual information* refers to the amount of information the representation provides. This is measured by the amount of aggregated gaze data used to generate the representation in addition to its overall size. We further sub-categorised the selected representations by their characteristics into three equally-sized subgroups: (1) Point-based, (2) Trajectory-based and (3) Area of Interest (AOI)-based.

[2]http://www.tobiipro.com

Representation	*Conspicuousness*	*Visual Information*	*Category*
Dot	Low	High	Point
Cursor	Low	Low	Point
Spotlight	Low	Moderate	Point
Fixation	High	Low	Trajectory
Scanpath	High	Moderate	Trajectory
Fixation Trail	Moderate	Moderate	Trajectory
Heatmap	Moderate	High	AOI
Convex Hull	Moderate	Low	AOI
Bee Swarm	High	High	AOI

Table 1: Real-time gaze representations in relation to conspicuousness and visual information.

Point-Based Representations

For this category, we first selected two types of gaze representations previously examined by Zhang et al. [29]: *cursor* and *spotlight*. We then added the commonly-used *dot* representation as a contrast to the first two in terms of precision and visual information. The point-based representations are mapped directly to the real-time gaze coordinates from the eye tracker. To reduce the jittery motions that naturally arise from eye movements, we applied a weighted moving average filter [27] on top of the eye tracker's already lightly filtered data to smooth out the motion. This gave all three representations a hovering effect when a user's eyes moved across the screen and reduced distracting jitter during gaze fixations.

Dot: This representation appears as a translucent red dot, akin to a laser pointer beam or, at larger sizes, a torch beam (Fig. 2a). It has commonly been adopted for use in gaze-based interaction as a deictic pointer and has been demonstrated to be an effective referential pointer to facilitate coordination in remote collaboration [1]. This representation allows a viewer to see gaze with high precision but can be potentially distracting. Our implementation used a 0.5 opacity with a radius of 30-px.

Cursor: This representation is a large ring with a transparent centre (Figure 2b). Its use has been demonstrated by Zhang et al. [29] and in GazeArchers [21], a game in which two players use gaze and touch simultaneously on a shared surface. Further, Tobii divisions has implemented a similar representation

in many of their recent products (Tobii Gaming[3] *Gaze Trace*, Tobii Pro *Gaze Share*[4]). Tobii Pro also uses this representation for their wearable eye tracker controller to view the gaze of a user in the real world in real-time. This representation is equivalent to the size of the foveal vision which is about the size of a thumb at arm's length, making explicit the confidence range of the gaze estimation. We matched the size of this representation to default size of Tobii's *Gaze Trace* (150-pixel radius) and coloured the outline in grey to give this a low visibility. This representation provides more visual information than the previous representation (*Dot*) as it allows users to see what is enclosed in the ring as opposed to a specific point.

Spotlight: Figure 2c depicts this representation where the gaze point is shown as a red circle with high opacity in the centre and decreasing opacity towards the edges. This was the most widely preferred representation in Zhang et al.'s evaluation [29]. It was noted to be more subtle and less distracting due to the reduced overall visibility and conspicuousness. We modified the parameters to suit our study, reducing its overall size to a radius of 90 pixels (30 pixels at the highest opacity) and changing the colour to red, as white blended into the light background of the map area of the *Ticket to Ride* game board.

Trajectory-Based Representations

The representations in this category use the sensitive fixation stream provided by the eye tracker's SDK to reveal the sequence of points at which the player has recently looked, leaving behind a short history of the player's gaze. We initially drew from professional eye tracking usability studies that use animated representations to replay gaze data. In our review, we found similar representations being used in eye tracking research. For example, Qvarfordt and Zhai use a scanpath representation to infer dynamic eye movements in their study [22]. We also found proposed representations aimed to enhance remote collaboration, demonstrating the importance of assessing representations that fall into this category [17]. A red fill is used for representations in this category, with each fixation presented as a translucent ellipse.

Fixation: This representation aims to illustrate points of focus by displaying a single, unmoving fixation point as a red dot (Figure 2d). When a player maintains their fixation, the dot remains small in size. If the distance to the next fixation is within a threshold, the dot becomes larger. In real-time, this representation gives the appearance of small eruptions occurring on the interface that indicates whether a player is sustaining attention in an area (large dots) or flicking around between multiple areas (small dots). This representation was inspired by the 'Live Viewer' feature in the Tobii software, which Higuch et al. implemented for use in a remote collaborative scenario [11]. Our adaptation removes the elements of direction (tail, fade) used in other implementations to potentially reduce the mental load. Therefore, participants see a single point that appears then disappears instead of following a point.

Scanpath: The scanpath representation shows the length, duration and direction of gaze by using a sequence of fixations connected by a line (Fig. 2e). Variation of this representation has long been used to visualise gaze patterns in both psychology, eye tracking usability studies and in HCI. This representation has been used unidirectionally such as in Qvarfordt and Zhai's study to show the interest of a remote unaware user, or bi-directionally such as in Jermann and Nussli to improve communication in a remote pair-programming scenario [13, 22]. We implemented this representation based on Goldberg and Helfman [10]. Many approaches have been presented to overcome the problem of visual clutter such as by bundling close fixations into a single point. Another approach is to use time duration such as in previous works (e.g. 5 seconds [28]. Instead, we opted to limit the number of fixation bundles displayed to 5, reduce the visibility of the lines connected the fixation bundle and removed other elements such the numbering of each fixation bundle to reduce visual information.

Fixation Trail: This representation was originally proposed by Li et al. to illustrate gaze paths as opposed to a single fixation point [17]. In our study, we renamed this representation to *Fixation Trail* (originally *Gaze Trail*) as it uses fixation data and leaves a trail of fixations (see Figure 2f). Li et al. believe that the availability of prolonged gaze information allowed participants to make references to multiple locations (e.g. contrasting two different fixation clusters). This representation is different to *Scanpath* in that we fixed the size of the fixation to a 30-px radius. Only 5 fixation points remain fully visible while the remainder fades over time.

AOI-Based Representations

In this category, we selected gaze representations that allow an observer to view the Areas of Interest (AOI) at which the user has looked as a whole rather than a specific point. Representations in this category use fixations aggregated over time. We can control how much visual information is displayed by adjusting its time window. Our implementation uses 120 fixation points ($\tilde{2}$ seconds) so as not occlude the screen (increase visibility, reduce visual information). We selected three representations with varying levels of visual information based on their popularity in eye tracking research.

Heatmap: Heatmaps represent the distribution of gaze points by a colour gradient (Figure 2g). We included this representation as it has commonly been used to visualise data in many domains and is not limited to eye tracking. In eye tracking usability studies it is also used for aggregating the gaze data from multiple participants. While the representation requires the aggregate participant data, it is possible to display a dynamic heatmap of a single tracked user [7]. Our implementation calculates the density of fixations within a 2-second window and displays the heatmap accordingly. The window size in our implementation attempts to balance visual information, allowing several 'hot spots' to be shown but do not remain for too long that it starts to occlude the interface.

Convex Hull: This representation is also known as *Cluster* (Tobii Pro) which display areas with a high concentration of gaze points as polygons to shows areas of interest (Figure 2h). Our implementation uses the fixation data stream in which we use to render a convex hull polygon in real-time using a clustering algorithm. As the stream continuously updates the convex hull, this creates a shifting polygon that shows the

[3] https://www.tobiigaming.com
[4] https://tobiicloud.com/share/

area in which the area is interested in. When the player looks rapidly between two points, the polygon stretches towards the points, but when the focus on a small area is high, the representation becomes small. This representation aims to provide minimal visual information of the area in which the participant has shown interest in.

Bee Swarm: Bee Swarm has been typically used to replay gaze points of several subjects against dynamic content such as videos, allowing efficient comparison of recordings from several subjects. Our variation aims to emulate this representation for use with a single user by displaying all fixation points in the last 2 seconds in which we enclose in a convex hull (Figure 2i). Each point was 5 pixels in radius with a red translucent fill. This visualization allows the player not only to see where the area in which the opponent was looking but the also points within that area. When the user focuses in a particular area, the points will become denser in the area, revealing an interest in the area. This representation has the highest amount of visual information as it combines characteristics from both AOI- and point-based representations.

RESEARCH DESIGN

In this study, we aim to understand the following research questions derived from our review of the literature:

[RQ1] How well are human subjects able to infer intent from a range of gaze representations?

[RQ2] What characteristics of gaze representations enable humans to perform better in predicting intentions?

[RQ3] Do gaze representations have an influence on predicting intent in a competitive gameplay setting?

To answer these questions, we adopted a spectator ('third person') approach. This means that participants other than the players watched clips of gameplay sessions after the fact, with different clips featuring different gaze representations. We chose to make our participants spectators rather than active players in order to simplify the task they were asked to perform, as they could focus on interpreting the gaze data without worrying about managing their own strategy. This decision was informed by the common finding in prior works that live gaze data can be highly distracting from a task at hand [17, 20, 29]. As this was the first study to assess people's ability to infer intent from gaze data of a remote opponent in a game setting, we decided to begin with a more concentrated test of interpretive ability, before introducing the distractions of live gameplay in a future study.

We recorded our data in two stages. In the first, we recruited 8 players to record 4 games of our selected game—*Ticket to Ride* (described below) in which one player's gaze was shared. In the second, a separate group of 27 participants watched short selected clips (< 1 minute) of the recorded gameplay from the perspective of the player who could see the real-time gaze representations. These participants provided an assessment of each clip, which we used in our analysis. A secondary benefit of this two-stage approach was that we were able to conduct a larger number of gaze visualization assessments in a shorter period of time than would have been possible if all participants had had to actively complete multiple games. This enabled a larger sample size for this foundational work, which we will build on in future studies involving active players.

Figure 3: Participant assessing a clip in a lounge room setting.

Application

We developed a custom networked system that overlays real-time gaze visualizations over any application (see through, click through). We used this application to stream gaze data between two laptops, each equipped with a Tobii EyeX eye tracker, in different rooms. We streamed two types of gaze data concurrently from the SDK: the lightly filtered gaze data and the sensitive fixation data. Our application is capable of displaying the gaze of one or both users simultaneously, meaning that we have the ability to switch on the gaze of one user to be visualised by themselves or by another user. Moreover, we can easily switch the gaze between representations through a control window on an extended screen.

Game and Stages

In *Ticket to Ride*, each player competes to claim train routes between cities across a map of North America, seeking to connect certain pairs of cities ('tickets') in a continuous path of linked routes. Only one player can claim each route, and players do not know which cities their opponent is seeking to connect. Therefore, players must place their routes carefully, staying alert to the risk that their opponent will claim a route before them and block the most convenient path between their goal cities. Keeping information hidden plays a large role in *Ticket to Ride*, as a player can gain significant advantages over their opponents by correctly guessing which routes they are planning to take. Hence, *Ticket to Ride* works as an excellent use case for our study for several reasons. First, it is a widely-played board game, already familiar to many, with a digital game version that is almost identical to the board game; and because it has relatively simple rules involving clear goals and actions, which can be learned in under 15 minutes by interested players who have not played it previously. Second, as a tabletop game, it demands constant attention, requiring regular thought processing and frequent changes to short- and long-term strategies. This affords us to obtain a range of behaviours to display. Third, Lankes et al. have determined that turn-based games are better-suited to gaze transfer applications (as opposed to real-time first person shooters) [16]. This is because such games give players the opportunity to alternate between playing their own moves and watching their opponents' behaviour as they take their turn. Fourth, it is a map-based game, in which the players' focus and strategic calculations are closely concentrated on the geography of the

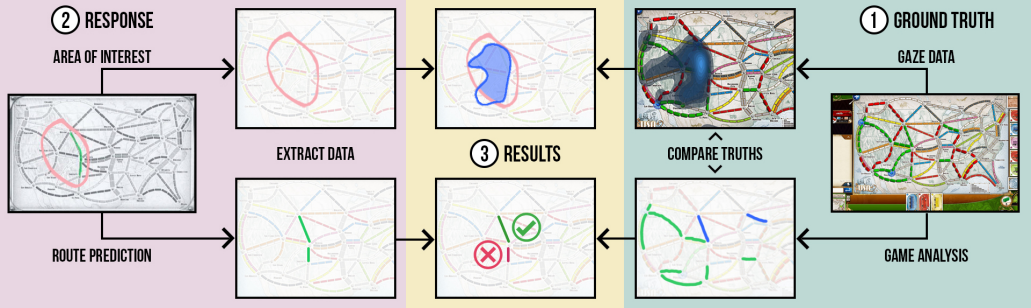


Figure 4: Method for analysis for Area of Interest (AOI) Estimation and Route Prediction.

map; gaze awareness and eye movement studies with map-based applications have been shown to be highly successful, as the static map gives users a point of reference [15, 22, 29].

As a playthrough of *Ticket to Ride* progresses, the map becomes increasingly populated with trains, and the strategic information available to the players becomes consequently richer while their choices of unclaimed routes become more constrained. Each player starts with a stock of 45 trains, and the game ends when any player's stock of unplaced trains falls below three. We use this train stock as a metric for game progression to divide the recorded games into three stages: early (45-31 trains), middle (31-16 trains), late (15-0 trains). The early stage is typically given over to planning, and placing trains are minimised as players try not to give away their goals. The middle stage sees greater competition for routes, and is where players often find themselves clashing over the most efficient paths between their respective goal cities. The late stage involves players either slotting in the last remaining routes that they need, seeking to block their opponent, or looking for other routes they can pick up extra points. We include an even selection of representations of each stage in our study, as we hypothesised that the usefulness and relevance of gaze visualizations might change as the game progresses.

Gaze Visualization Recording

We recruited 8 players (4 pairs) to record the necessary gameplay clips between a *naive player* (whose gaze was being recorded) and an *aware player* (for whom their opponent's gaze was visualised). Both players were placed in separate rooms with an experimenter and given the opportunity to play the in-game tutorial for up to 10 minutes to understand the rules and mechanics of the game. Once the *naive player* was confident with the game, they were asked to perform the default calibration provided by the eye tracker. The *naive player* does not know that they are being watched to record and observe 'natural gaze'. Meanwhile, the *aware player* was briefed that different gaze visualizations would be shown to them and that their opponent was not aware that their gaze was being shown. Each pair played three games, and the gaze representation was changed at every stage of the game. We recorded the displays using *Open Broadcaster Software (OBS) Studio*[5] and logged the gaze data with our custom application.

[5]https://obsproject.com/download

Following the recordings, we reviewed the videos and noted instances in which the naive player exhibited various behaviours, such as planning, checking game counters or tickets, and looking for routes to claim on the map. From the videos, we selected three short clips for each of the nine gaze visualization formats: one in early game, one in middle game and one in late game. This gave us a total of 27 clips, exhibiting a variety of behaviours, and balanced between stages to control for the influence of game stage. Moreover, as different visualizations were used in different stages of different games, we are able to randomise the clips further. The length of the clips ranged between 20 and 60 seconds. All clips were saved without audio, to remove any potential sound clues such as the recorded player speaking aloud about their opponents' plans. As expected, the *aware players* experienced a level of distraction causing them to be drawn away from their own strategy and that some visualizations were preferred over others. However, we disregarded the feedback to avoid any early speculations.

STUDY

For the second stage of our study, we recruited 27 participants from the University of Melbourne. There were 11 male and 16 female participants, aged 18 to 44 years (mean=25.6), and included undergraduate students, graduate students and non-faculty staff. Only two had previously played the board game version of *Ticket to Ride*, but 17 indicated they had played a game of the same type in the past week. Participants were compensated with a $5 coffee voucher for their time. We encouraged participants to watch a short video[6] that describes the rules and objectives as part of the recruitment process.

Procedure

Upon arrival, participants were greeted and seated comfortably in a lounge room setting, facing a 48-inch (1080p) television display connected to a PC as depicted in Figure 3. They completed a consent form and a simple demographic questionnaire and were briefed on the rules of *Ticket to Ride*. The facilitator showed them an example clip of gameplay footage and with it, explained how they should assess each clip on a paper survey, with the measures outlined under 'Measures and Analysis'. Here, the facilitator ensured that the participants understood the rules of the game and what was required of them before starting the study. Each session ran between 30 to 60 minutes.

[6]https://www.youtube.com/watch?v=uDODCHcj6Sg

Representation	*Early*	*Mid*	*Late*	*Overall*
Heatmap	.44(.78)	.68(.63)	.73 (.75)	.62(.72)
Fixation Trail	.48(.41)	.66(.72)	.48 (.75)	.54(.62)
Dot	.59(.49)	.37(.55)	.67(.36)	.54(.47)
Cursor	.46(.73)	.68(.68)	.44(.28)	.53(.56)
Fixation	.63(.72)	51(.87)	.41(.80)	.52(.80)
Scanpath	.40(.65)	.50(.79)	.65(.70)	.52(.71)
Bee Swarm	.59(.58)	.36(.63)	.43(.72)	.46(.64)
Convex Hull	.40(.62)	.17(.44)	.23 (.44)	.27(.50)
Spotlight	.24(.37)	.25(.43)	.28 (.65)	.26(.48)

Table 2: AOI Estimation Results. Precision(Recall).

Representation	*Early*	*Mid*	*Late*	*Overall*
Bee Swarm	.70(.37)	.67(.43)	.43 (.28)	.60(.36)
Cursor	.69(.85)	.54(.37)	.53 (.19)	.59(.47)
Heatmap	.57(.31)	.63(.29)	.50(.63)	.57(.41)
Scanpath	.40(.46)	.45(.26)	.63(.25)	.50(.32)
Spotlight	.50(.22)	49(.26)	.25(.14)	.41(.20)
Fixation Trail	.44(.14)	.50(.17)	.26(.26)	.40(.19)
Dot	.59(.37)	.37(.14)	.10(.04)	.35(.18)
Fixation	.37(.44)	.26(.14)	.27(.21)	.30(.26)
Convex Hull	.25(.11)	.27(.10)	.74(.02)	.20(.08)

Table 3: Route Prediction Results. Precision(Recall).

We employed a repeated-measures design: each participant was shown all 27 (3 stages $\times$ 9 visualizations) clips, each of which was considered one study 'condition', representing one gaze visualization format and one game stage. Participants were divided into three sets (A, B and C), and the order of the clips was randomised for each set.

Participants were given a maximum of 10 seconds to analyse the game and board state before the clip was played. In almost all cases, participants indicated that this was enough time for them to understand the game state, either with a verbal acknowledgement or a non-verbal indication such as a nod of the head. This time limit was imposed to prevent the participant from basing their inferences too heavily on reasoning about the layout of the game board, rather than the visualizations of the gaze data, and in order to more closely approximate a live gameplay setting in which the gaze information would be processed and responded to quickly. The facilitator encouraged participants to *think aloud* about what they were seeing, and to ask for any clarifications about the game if required. After the participant had viewed and assessed all 27 clips, they were finally asked to rank the nine gaze visualization formats from most preferred to least preferred. In summary, we collected a total of 27 participants $\times$ 9 visualizations $\times$ 3 stages $= 729$ responses for each of our dependent measures.

Measures and Analysis

In response to each clip, participants were asked to record the following five pieces of information: (1) *Area of Interest Estimation*, using a map of the game board they circled the area (or multiple areas) in which they believed the gaze-tracked player had been interested. (2) *Route Prediction*, on the same map they highlighted any specific routes that they thought the gaze-tracked player was planning to build upon. They were also encouraged to write a short interpretation of what the gaze-tracked player was thinking during the clip or to explain their interpretation of the gaze behaviour out. (3) They rated how informative they found the gaze visualization in the clip on a 7-point scale ranging from *Not informative at all* to *Extremely informative*. (4) They rated how well they felt they could predict the gaze-tracked player's intentions in the clip on a 10-point rating scale. Participants were permitted to skip any of these questions when they felt unable to give a response, for example, if they could not predict any routes upon which the gaze-tracked player was considering to build. Finally, we collected (5) *Qualitative Responses* at the end of the session, participants were asked to rank the representations in order of preference and explain the reasoning behind their choices.

Area of Interest Estimation and Route Prediction

To determine how closely each participant guessed the area of interest of the naive player, we compared the intersection of each response against the ground truth for the respective condition. Participants recorded their responses by circling the corresponding area in the map. To be able to analyse this data digitally, we scanned the highlighted pages and extracted the contours of the highlights using a custom-built OpenCV computer vision application. To establish a ground truth, we generated 2D density maps from the gaze data in each clip using the R statistical package. We then refined the AOIs against the manual analysis of the clips, discarding outliers that did not correspond to the routes the player eventually built. We then compared the intersection of the two regions, computing the precision (area of intersection divided by the total area circled by the participant) and the recall (area of the intersection divided by the area of the ground truth contour) of the participant's estimate (see Figure 4, top path).

Similarly to the analysis of the AOI Estimation, we scanned the highlighted maps and extracted the highlighted routes using our OpenCV application (see Figure 4, bottom path). We established the ground truth by recording the routes that the player eventually built after the clip. We also recorded routes that the player was considering to build but never did. We inferred such routes based on evidence from their hidden goal cards, subsequent game actions, and in-game comments. As with the AOI Estimation, we operationalise their performance primarily based on precision and recall, but also considered their harmonic mean i.e. *F-Score* ($F = 2 \times (precision \times recall)/(precision + recall)$). The details of these performance metrics are explained by Rijsbergen [23].

Quantitative and Qualitative Responses

For our third and fourth measure, we will use descriptive statistics to analyse how informative as well as the predictability of the gaze representation. To gain richer insights beyond the quantitative performance data, we collected subjective perceptions in three ways. First, we transcribed all verbal comments as they thought out loud when watching the clips. Second, in the questionnaires for the clips, participants answered an open-ended question that asked them to predict what the player was thinking. Third, we transcribed the interview about their order of preference at the end of the study. We were particularly interested in comments beyond the areas of interest and routes, including potential strategies, such as building a route to block the opponent, planning a route, verifying whether they have enough resources to build a certain route.

Representation	*Early*	*Mid*	*Late*	*Overall*
Spotlight	5.1 (1.0)	5.1 (1.1)	4.8 (1.0)	5.0 (1.0)
Heatmap	4.8 (1.0)	5.0 (0.8)	4.9 (1.0)	4.9 (0.9)
Cursor	4.9 (1.3)	4.8 (1.4)	4.6 (1.3)	4.8 (1.3)
Dot	5.1 (1.6)	4.8 (1.1)	4.1 (1.7)	4.7 (1.5)
Fixation	4.7 (1.0)	4.5 (1.2)	4.6 (1.1)	4.6 (1.1)
Scanpath	4.7 (1.0)	4.5 (1.2)	4.8 (1.0)	4.6 (1.1)
Bee Swarm	4.6 (1.1)	4.6 (1.3)	4.4 (1.4)	4.5 (1.3)
Fixation Trail	4.4 (1.2)	4.7 (1.0)	4.3 (1.0)	4.5 (1.1)
Convex Hull	3.2 (1.2)	3.6 (1.4)	3.0 (1.5)	3.3 (1.4)

Table 4: Ratings of Informativeness Results 1 Not Informative at all - 7 Extremely Informative). Score (Standard Deviation)

RESULTS

We report the results of the nine visualizations at each of the three game stages and their overalls. Five primary metrics are reported: the percentage overlap of the AOI indicated by the participant against the AOI detected by the software; the performance between routes predicted by the participant and routes plausibly considered by the naive player; participant ratings of the visualizations' informativeness; participant self-assessments of how well they can predict what the naive player is planning to do; and participant rankings of gaze visualization preference. Lastly, we compared the scores from their performance in our first two metrics against the last three.

Area of Interest Estimation

Table 2 shows the precision and recall for each representation, ordered by precision. The order of the performance of the visualizations in terms of precision were the *Heatmap* (.62), followed by the *Fixation Trail* (.54), and finally the *Dot* (.54). This means that these are the best representations for selecting areas of interest within the ground truth. In terms of recall, the *Fixation* (.80) was the best performing, followed by the *Heatmap* (.72) and the *Scanpath* (.71). This indicates that these were the representations in which the selected area captured a larger area within the area of interest. If we combine both metrics using the *F-Score*, we find that the overall best representation was the *Heatmap* (.66), followed by the *Fixation* (.63) and the *Scanpath* (.63). To test whether the stage of the game had an effect on the precision and recall, we conducted a repeated-measures ANOVA for each representation. Except for *Spotlight*, we found that the stage had a significant effect on the stage on the precision and recall. Precision significantly increased in the *Heatmap* and *Fixation Trail* as the game progressed, and decreased in the remaining conditions.

Route Prediction

Table 3 shows the computed precision and recall results for each visualization ordered by precision. The best performing visualizations when it comes to route prediction in terms of *F-Score* are *Cursor* (.52), *Heatmap* (.48), and *Bee Swarm* (.45). *Convex Hull* was ranked significantly lower in both precision and recall (.20 and .08 respectively) and this was primarily from the late stage precision score (.02). The other low score noted, also in the late stage, is the *Dot* representation recall score (.04). We tested the effects of the gameplay stage on both precision and recall for each representation using repeated-measures ANOVA. For precision, only the *Scanpath* representation showed significance ($p = 0.014$) in which a

Representation	*Early*	*Mid*	*Late*	*Overall*
Heatmap	6.3 (2.0)	6.7 (1.9)	7.1 (1.7)	6.7 (1.9)
Cursor	7.0 (1.7)	6.1 (2.0)	6.6 (2.0)	6.6 (1.9)
Scanpath	6.0 (1.5)	6.0 (2.2)	6.6 (1.6)	6.2 (1.8)
Fixation	6.9 (1.7)	5.4 (2.1)	6.3 (1.5)	6.2 (1.9)
Spotlight	6.2 (1.7)	6.5 (1.5)	5.6 (1.9)	6.1 (1.8)
Bee Swarm	5.9 (2.1)	6.2 (2.0)	5.8 (2.3)	6.0 (2.1)
Fixation Trail	6.1 (1.7)	6.2 (1.5)	5.6 (2.2)	6.0 (2.0)
Dot	6.9 (2.3)	5.7 (1.8)	3.8 (2.0)	5.5 (2.4)
Convex Hull	4.0 (1.8)	4.7 (2.0)	3.9 (2.1)	4.2 (2.0)

Table 5: Predictive Ability Results. Score (Standard Deviation)

Bonferroni post hoc test showed a significant difference between the mid and late stage ($p = 0.0069$). Here, we note that the *Scanpath* representation is the only representation to have increase in precision from the mid stage ($M = 45.5$) to the late stage ($M = 63.4$). A majority of representations, 6 out of 9 (*Bee Swarm, Heatmap, Spotlight, Fixation Trail, Scanpath and Convex Hull*), remained precise for the early and mid stages but significantly becomes less precise in the late stage. Both *Fixation* and *Cursor* precision dropped significantly from early ($M = 37.3, M = 69.1$) to mid game ($M = 27.2, M = 54.0$) but remained consistent in the late game. The precision of the *Dot* representation dropped linearly from early to late stage. There was no significant effect on recall on the gameplay stages.

Ratings of Informativeness

The responses follows a normal distribution (*Pearson chi-square normality test,* $p < 0.05$), in which calculated the mean scores and standard deviation for each visualization for each stage, and overalls (see Table 4). We test the effects of stage on their rating on informativeness using repeated-measures ANOVA and found a significance in the mid and late stage ($p = 0.0247$). We tabulated the results and ranked them by overall rating accordingly. We found that *Spotlight* (5.0), *Heatmap* (4.9) and *Cursor* (4.8) received the highest rating while *Convex Hull* was rated significantly lower (3.3). *Heatmap* had the lowest deviation (0.9) among the visualizations showing an agreement among participants on its informativeness.

Self-Assessment of Predictive Ability

We ran the same statistical tests used in the previous measure and found that the responses also follows a normal distribution. We calculated the mean scores and standard deviation of each visualization for each stage and overalls, and ranked them by overall rating (see Table 5). We proceeded to test the effects of stage on the responses using repeated-measures ANOVA and found significance in the early and late stage ($p = 0.331$). The results shows that *Heatmap* (6.7), *Cursor* (6.6) and *Scanpath* (6.2) received the highest rating while *Dot* (5.5) and *Convex Hull* (4.2) was rated significantly lower than the others.

Preference Rankings

Figure 5 shows the ranking data for each representation by all participants, showing that the *Convex Hull* was ranked the lowest ($M = 2.0$) while *Heatmap* was the highest ranked ($M = 6.7$). *Cursor* and *Scanpath* also was ranked highly. We further analysed the ranked data using the Friedman test in which we found a significance ($p < 0.05$). The post hoc test showed that *Convex Hull* representation was ranked significantly lower than

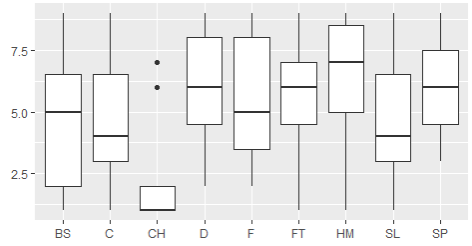


Figure 5: Boxplot of Preference Ratings given for each visualization. (BS) Bee Swarm, (C) Circle, (CH) Convex Hull, (D) Dot, (F) Fixation, (FT) Fixation Trail, (HM) Heatmap, (SL) Spotlight and (SP) Scanpath.

all other representations. We ran the Friedman test once again but this time removing the *Convex Hull* representation from the test. We found that there were no significance differences between the other rest ($p = 0.07547$). We proceeded to analyse the interviews to gain further insights into these preferences.

We received many positive insights into why *Heatmap* was ranked the highest and preferred by a majority of participants. Participants that gave the *Heatmap* a high score mentioned its 'persistence' i.e. the visualization gave them a moment to make an inference before it fades away (e.g. *"I like that it stayed on the screen for awhile" [P3]*). A small number of participants (e.g. P7) noted that they were familiar with the visualization having seen or used them in other domains, noting its inherent usefulness. When participants were asked why they preferred *Heatmap* over the others, a majority they mentioned that they had difficulty following them. Two contributing factors for this were (1) speed (*"too quick" [P16]*, *"moved too much and too quickly"* [P11]) and (2) lack of conspicuousness (e.g. *"more subtle ones are harder to keep track of" [P19]*, *"it was hard to see" [P25]*). Some participants mentioned that the *Scanpath* visualization blended into the game making it hard to differentiate (e.g. P20) while others found it useful (e.g. *"I like it because it gives you a track." [P11]*). Conversely, some participants did rate visualization that falls under point-based category highly. Participant 22 provided his explanation, stating that he preferred the point-based representations as he was able to process the visual information quickly and described trajectory- or AOI-based representations as "a lot going on". Participants often compared *Convex Hull* and *Bee Swarm* and found that the 'red dots' made it better (e.g. P17). *Bee Swarm*'s ability to have persistence yet show movement at the same time it is probably a reason why it scored well in the route prediction measure. Overall, participants provided rich feedback, noting that AOI-based representations were good for predicting overall strategy while point-based representations appear to be better moment-to-moment attention.

DISCUSSION

Our study explored the characteristics of nine visual representations of gaze to determine which was most effective at conveying intent to an observer by exploring relationship between the characteristics of representations and human ability to infer intent from them. This section discusses the results from the five metrics as well as the post study interviews.

Representation Performance and Preference

Once again *Convex Hull* was largely indecipherable to participants, with the lowest scores for route prediction and close to the lowest scores for AOI estimation. However, the other AOI-based visualizations, *Bee Swarm* and *Heatmap*, had among the best results for route prediction. Unlike *Convex Hull*, these visualizations can clearly display both the broad area of sustained interest and the focal point of the player's moment-to-moment attention. The *Heatmap* shows multiple hot spots and recent zones in which the player has looked. Interestingly, *Bee Swarm* performed well for route prediction but relatively poorly in AOI estimation measure. Our version of the *Bee Swarm* is encompassed by a convex hull display, which allows viewers to see at a glance the general region of the player's gaze while their more exact moment-to-moment attention is captured by a large number of rapidly-propagating red dots. We believe that participants used the convex hull to narrow their focus, enabling them to see the routes more clearly; avoiding a visual search for the next sequence.

Conversely, *Fixation Trail* and *Dot* performed strongly for AOI estimation, but relatively poor for route prediction, suggesting that these visualizations were effective at conveying general information but were not fine-grained enough for inferring specific intent. We believe that trajectory-based representations caused participants to divide their attention between two areas of interest and this was specifically mentioned by Participant 22. None of the point-based representations performed particularly well. We believe by the time participants start to make a connection between objects, the visual has moved somewhere else, causing them to start making a new inference elsewhere.

The *Heatmap* gaze representation overall scored highly in all five measures, and this is primarily attributed to its ability to provide a high level of visual information by using averaged information. Moreover, *Heatmap* takes a small amount of time to 'heat up' and cool down, allowing the player to know areas which are gaining interest and losing interest. We earlier noted that *Scanpath* performance increased from the early-mid to late stage of the game which may be reflective of the nature of the game. Towards the late stage players tend to track the missing paths which reduces the search space and enhances the performance of the *Scanpath* gaze representation. Further, with the increased availability of other cues to select from as the game progresses, a sustained gaze (presenting sustained interest) at a location is likely to signal with greater confidence the likelihood of an intention to action. To some degree, we expected the *Heatmap* to have performed the best being a well-used representation. On the other hand, the *Convex Hull* was notably disliked, being given the lowest or second-lowest ranking by all but four participants.

Preferences were mixed for the rest, with a wide variation in scores indicating that different participants valued different characteristics. The trajectory-based representations were all ranked relatively highly; we hypothesise that this is because trajectory-based representations naturally calls attention to the connections between different points of interest as the player moves their gaze between them without overloading the user and therefore remained neutral in our results.

Interplay Between Game Cues and Overall Scores

As expected, we found that participants make use of in-game and behavioural cues to predict routes such as by observing what cards the naive player looked at or picked. From the game analysis, there was 17 showed cards being picked and many participants drew conclusions by matching the route seen and card colour which is a probable cause of differences in scores between the clips. Across all visualizations, prediction accuracy was lower in the late game than in the early and mid game. We consider several factors that may account for this trend, with different implications for the use of gaze visualizations in other contexts. First, the visual clutter on the screen increases as the game progresses, so that in the late game there are substantially more train routes that may potentially distract the viewer from the gaze representation. Second, the number of available routes is lower in the late game, which tends to require players to draw more convoluted routes; the combination of zigzagging existing routes and fewer possibilities may increase the difficulty of recognising an obvious path that the player is aiming for. Third, the behaviour of the player is likely to be different, as they transition from long-term planning in the earlier stages to opportunistic route-grabbing and path-blocking in the late stage; this behaviour may be more difficult for a viewer to interpret. We believe that it is plausible that all three of these factors play a role, but based on our own observation of the game recordings, we expect that the third factor to be the most significant as gaze behaviour clearly shows a different pattern late in the game. This is exhibited in our results such as in the case of the *Dot* representation where the naive player was looking for was considering a route to build between a few potential routes.

Humans naturally rely on explicit actions as they confirm their thoughts, such as in real life. Regardless of representation shown, we are impressed at the ability of participants to put themselves into the mind of the player with the introduction of displayed gaze. In the same way, spectators rely on these actions, and therefore participants employ a variety of cues in making inferences. However, we observed in our study that some participants better than other in keeping track of the elements in the game, allowing them to make better inferences.

Design Implications

Our results present various implications towards the digital games community. First and foremost, we demonstrated the possibilities of using different types of representations to display mental processes (e.g. intention) of players in real-time, which has direct implications towards *game developers*. When using gaze explicitly for use in games, game developers must be wary of how it is actually used in its execution. For example, when we the tested the *Convex Hull* representation, it worked well allowing two developers to reveal what the other was looking at. However, during our evaluation, *Convex Hull* obtained consistently low scores revealing a flaw in our implementation when converting a static visualisation to real-time. Then again, this consistently low score gave us a reference point and confidence in the validity of our data. The design of gaze representations should further consider the amount it reveals without giving it away as it would give a player a disadvantaged. Imagine playing a game and your opponent is constantly one step ahead of you. For the evaluation of games in real-time with gaze visualisation (implicit use), representations such as the *Heatmap* can be used to reveal intentions of players from natural gaze behaviour, and therefore provide insights into whether players understand the task in hand or is making correct strategic decisions. Second, a range of selectable gaze visualizations can implemented in live game streaming platforms such as Twitch, where *audiences* to create sensory excitement and anticipation. For example, audiences can anticipate intentions of live players before the actual move, providing a new experience. Lastly, *players* themselves can benefit from visualisations the gaze of their opponent in the form of an in-game 'power up' to create new gameplay experiences, especially in similar competitive contexts.

CONTRIBUTIONS, LIMITATIONS AND FUTURE WORK

This work presents three main contributions. First, we provide an understanding of the role of gaze and its importance to game designers in choosing and designing the right visualisation. Second, an assessment of how people interpret visual gaze visualisations, which has not been done before in this context. The findings also highlight the effect of the different techniques on the inferring ability. Third, we contribute the assessment of AOI-based representations in the context of gameplay. Limitations wise, we used clips in our study from the entire game session which might not be comprehensive to show the goals of the whole game but enough to determine short term intention. Moreover, as each visualization was tested in only three different clips, it is possible that one outlier clip could affect the overall result for that visualization; nevertheless, we have taken this into account in our reporting of the findings. We also acknowledged that we only used an example of a strategic game to demonstrate gaze. However, we believe our findings apply to similar games.

Our future work focuses on *players*, where we give one player the ability to predict intentions using a real-time *Heatmap*, and with room for deceptive or misinterpreted behaviours to happen. We have also considered testing the viability of our findings with a different game (e.g. on with a less cluttered board) to provide further insights into the generalisation of our results. Furthermore, our work has further implications in gaze-aware artificial intelligence, and we have begun putting together a planning-based AI that predicts players potential intentions based on the recorded games from this study.

CONCLUSION

This paper contributes the groundwork for determining the characteristics of gaze representations that allow human players to make inferences about other players' intentions and to derive insights into potential strategies. This has been achieved by measuring the performance of human in their ability to predict intentions using a range of visual gaze representations. Our results showed us that human subjects have a remarkable ability to infer intent with the introduction of gaze. We conclude that representations that show both characteristics of point- and AOI- based representations would work well in our setting. Our work ultimately contributes gaze input as a viable implicit modality in multiplayer competitive computer games and with broader implications towards the games community.

ACKNOWLEDGEMENTS

We acknowledge the Australian Commonwealth Government and the Microsoft Research Centre for Social NUI for their support on this project. We further acknowledge Zaher Joukhadar for major improvements made on the custom built software.

REFERENCES

1. Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1151–1158. DOI: http://dx.doi.org/10.1145/2851581.2892459
2. T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. 2014. State-of-the-Art of Visualization for Eye Tracking Data. In *EuroVis - STARs*, R. Borgo, R. Maciejewski, and I. Viola (Eds.). The Eurographics Association. DOI: http://dx.doi.org/10.2312/eurovisstar.20141173
3. Susan E. Brennan, Xin Chen, Christopher A. Dickinson, Mark B. Neider, and Gregory J. Zelinsky. 2008. Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition* 106, 3 (2008), 1465 – 1477. DOI: http://dx.doi.org/10.1016/j.cognition.2007.05.012
4. Marcus Carter, Joshua Newn, Eduardo Velloso, and Frank Vetere. 2015. Remote Gaze and Gesture Tracking on the Microsoft Kinect: Investigating the Role of Feedback. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction (OzCHI '15)*. ACM, New York, NY, USA, 167–176. DOI: http://dx.doi.org/10.1145/2838739.2838778
5. Mauro Cherubini, Marc-Antoine Nüssli, and Pierre Dillenbourg. 2008. Deixis and Gaze in Collaborative Work at a Distance (over a Shared Map): A Computational Model to Detect Misunderstandings. In *Proceedings of the 2008 Symposium on Eye Tracking Research & Applications (ETRA '08)*. ACM, New York, NY, USA, 173–180. DOI: http://dx.doi.org/10.1145/1344471.1344515
6. Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2492–2496. DOI: http://dx.doi.org/10.1145/2858036.2858499
7. Andrew T. Duchowski, Margaux M. Price, Miriah Meyer, and Pilar Orero. 2012. Aggregate Gaze Visualization with Real-time Heatmaps. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 13–20. DOI: http://dx.doi.org/10.1145/2168556.2168558
8. Paul Ekman and Wallace V. Friesen. 1969. Nonverbal Leakage and Clues to Deception. *Psychiatry* 32, 1 (1969), 88–106. DOI: http://dx.doi.org/10.1080/00332747.1969.11023575
9. Alexandra Frischen, Andrew P Bayliss, and Steven P Tipper. 2007. Gaze cueing of attention: visual attention, social cognition, and individual differences. *Psychological Bulletin* 133, 4 (2007), 694.
10. Joseph H. Goldberg and Jonathan I. Helfman. 2010. Visual Scanpath Representation. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (ETRA '10)*. ACM, New York, NY, USA, 203–210. DOI: http://dx.doi.org/10.1145/1743666.1743717
11. Keita Higuch, Ryo Yonetani, and Yoichi Sato. 2016. 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 (CHI '16)*. ACM, New York, NY, USA, 5180–5190. DOI: http://dx.doi.org/10.1145/2858036.2858438
12. Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-computer Interaction Techniques: What You Look at is What You Get. *ACM Trans. Inf. Syst.* 9, 2 (April 1991), 152–169. DOI: http://dx.doi.org/10.1145/123078.128728
13. Patrick Jermann and Marc-Antoine Nüssli. 2012. Effects of Sharing Text Selections on Gaze Cross-recurrence and Interaction Quality in a Pair Programming Task. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 1125–1134. DOI: http://dx.doi.org/10.1145/2145204.2145371
14. Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26 (1967), 22 – 63. DOI:http://dx.doi.org/10.1016/0001-6918(67)90005-4
15. Peter Kiefer, Ioannis Giannopoulos, and Martin Raubal. 2013. Using Eye Movements to Recognize Activities on Cartographic Maps. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL'13)*. ACM, New York, NY, USA, 488–491. DOI: http://dx.doi.org/10.1145/2525314.2525467
16. Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. 2016. 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 (ACE2016)*. ACM, New York, NY, USA, Article 17, 9 pages. DOI: http://dx.doi.org/10.1145/3001773.3001774
17. Jerry Li, Mia Manavalan, Sarah D'Angelo, and Darren Gergle. 2016. Designing Shared Gaze Awareness for Remote Collaboration. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, 325–328. DOI: http://dx.doi.org/10.1145/2818052.2869097

18. Paivi Majaranta, Hirotaka Aoki, Mick Donegan, Dan Witzner Hansen, and John Paulin Hansen. 2011. *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies* (1st ed.). Information Science Reference - Imprint of: IGI Publishing, Hershey, PA.

19. Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. 2016. *EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game*. Springer International Publishing, Cham, 102–114. DOI: http://dx.doi.org/10.1007/978-3-319-46100-7_9

20. Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. 2016. Exploring the Effects of Gaze Awareness on Multiplayer Gameplay. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY Companion '16)*. ACM, New York, NY, USA, 239–244. DOI:http://dx.doi.org/10.1145/2968120.2987740

21. Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. GazeArchers: Playing with Individual and Shared Attention in a Two-player Look&Shoot Tabletop Game. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 213–216. DOI: http://dx.doi.org/10.1145/3012709.3012717

22. Pernilla Qvarfordt and Shumin Zhai. 2005. Conversing with the User Based on Eye-gaze Patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*. ACM, New York, NY, USA, 221–230. DOI: http://dx.doi.org/10.1145/1054972.1055004

23. C. J. Van Rijsbergen. 1979. *Information Retrieval* (2nd ed.). Butterworth-Heinemann, Newton, MA, USA.

24. Randy Stein and Susan E. Brennan. 2004. Another Person's Eye Gaze As a Cue in Solving Programming Problems. In *Proceedings of the 6th International Conference on Multimodal Interfaces (ICMI '04)*. ACM, New York, NY, USA, 9–15. DOI: http://dx.doi.org/10.1145/1027933.1027936

25. Jayson Turner, Eduardo Velloso, Hans Gellersen, and Veronica Sundstedt. 2014. EyePlay: Applications for Gaze in Games. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play (CHI PLAY '14)*. ACM, New York, NY, USA, 465–468. DOI: http://dx.doi.org/10.1145/2658537.2659016

26. Eduardo Velloso and Marcus Carter. 2016. The Emergence of EyePlay: A Survey of Eye Interaction in Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, New York, NY, USA, 171–185. DOI: http://dx.doi.org/10.1145/2967934.2968084

27. Oleg Špakov. 2012. Comparison of Eye Movement Filters Used in HCI. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 281–284. DOI: http://dx.doi.org/10.1145/2168556.2168616

28. Nadir Weibel, Adam Fouse, Colleen Emmenegger, Sara Kimmich, and Edwin Hutchins. 2012. Let's Look at the Cockpit: Exploring Mobile Eye-tracking for Observational Research on the Flight Deck. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA '12)*. ACM, New York, NY, USA, 107–114. DOI: http://dx.doi.org/10.1145/2168556.2168573

29. Yanxia Zhang, Ken Pfeuffer, Ming Ki Chong, Jason Alexander, Andreas Bulling, and Hans Gellersen. 2017. Look Together: Using Gaze for Assisting Co-located Collaborative Search. *Personal Ubiquitous Comput.* 21, 1 (Feb. 2017), 173–186. DOI: http://dx.doi.org/10.1007/s00779-016-0969-x

A.2 Paper: *Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay*

DOI: 10.1145/3173574.31738357

Looks Can Be Deceiving: Using Gaze Visualisation to Predict and Mislead Opponents in Strategic Gameplay

Joshua Newn, Fraser Allison, Eduardo Velloso, Frank Vetere
Microsoft Research Centre for Social NUI
School of Computing and Information Systems, The University of Melbourne
{joshua.newn | fallison | evelloso | f.vetere}@unimelb.edu.au

ABSTRACT
In competitive co-located gameplay, players use their opponents' gaze to make predictions about their plans while simultaneously managing their own gaze to avoid giving away their plans. This socially competitive dimension is lacking in most online games, where players are out of sight of each other. We conducted a lab study using a strategic online game; finding that (1) players are better at discerning their opponent's plans when shown a live visualisation of the opponent's gaze, and (2) players who are aware that their gaze is tracked will manipulate their gaze to keep their intentions hidden. We describe the strategies that players employed, to various degrees of success, to deceive their opponent through their gaze behaviour. This gaze-based deception adds an effortful and challenging aspect to the competition. Lastly, we discuss the various implications of our findings and its applicability for future game design.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces: Input devices and strategies

Author Keywords
Deception; Gaze Visualisation; Competitive Gameplay; Intent Recognition; Plan Recognition; Nonverbal Leakage

INTRODUCTION

In face-to-face communication, our eyes have a dual function: to perceive the environment and to signal our point of interest within that environment [20]. In this context, gaze is an important behavioural cue that reveals our unspoken intentions to other people [7]; researchers call this "nonverbal leakage" [15]. As a result, interpreting gaze is a competitive element in many games, where players can gain advance warning of what their opponent intends to do by accurately reading their gaze. Variations on this tactic abound in all types of games, from poker players wearing sunglasses indoors to hide their eyes, to footballers using a "no look pass" to mislead opponents about the direction in which they will move the ball. As these examples show, counter-strategies for gaze interpretation range from the defensive (obscuring gaze) to the delusive (portraying a false intention through gaze).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI 2018, April 21–26, 2018, Montreal, QC, Canada
© 2018 ACM. ISBN 978-1-4503-5620-6/18/04…$15.00
DOI: https://doi.org/10.1145/3173574.3173835

By contrast, digital games rarely allow for gaze interpretation, as players are hidden behind screens and sometimes physically remote from each other. In a recent survey of gaze interaction in games, Velloso and Carter found an increasing number of applications in single-player scenarios [41], but limited usage in multiplayer and online games. HCI researchers have begun to explore the uses of shared gaze awareness in multiplayer games, finding that gaze visualisation can increase the feeling of social presence between remote players [27, 33]. However, despite its prevalence in face-to-face games, *gaze-based deception* in online gameplay has not yet been the subject of published research. This paper contributes the first study (to our knowledge) of human-to-human deception through gaze visualisation in an online strategic game.

In this study, we observed pairs of players engaged in an online turn-based strategy game with incomplete information, in the form of hidden objectives. Each player's gaze was tracked and, in some conditions, transmitted to their opponent and displayed as a dynamic heatmap overlaid on the game. The study involved three rounds of the game, corresponding to three study conditions: *no gaze visualisation*, *gaze visualisation of an unaware player*, and *gaze visualisation of an aware player*. First, consistent with the findings of our prior study [34], the findings of this study showed that players made more comprehensive and more precise predictions about their opponent's intentions when they were shown a gaze visualisation of an opponent that was unaware they were being tracked. Second, we find that players who knew their own gaze was being visualised were able to deceive their opponent with gaze, through both obscuring their true intentions and conveying false intentions. Following, we discuss the implications and applicability of these findings for future game design. Lastly, this paper contributes a typology of such gaze deception behaviours, categorised as *dissimulation* and *simulation deception strategies* according to the general theory of deception [4, 5].

LITERATURE REVIEW

Gaze and Intention

Gaze is a powerful form of nonverbal communication, as its direction alone can convey a number of signals, including one's current point of attention and future intentions [26]. Gaze often gives away this private information without our knowledge or intent, as "nonverbal leakage" [15]. It is common therefore for players in face-to-face games to peek at their opponents' eyes to obtain clues as to what they will do next. Psychologists theorise that this inherent human ability to discern others' intentions stems from our *Theory of Mind*—the ability to put

ourselves into another person's position and imagine how they might think and act—and that gaze is an important component to this ability [3, 28].

Gaze observation is a missing component in much technology-mediated human-human interaction. With the emergence of affordable eye trackers, researchers have begun to explore its uses in remote collaborative settings using visual representations of gaze (e.g. [2, 10, 30, 40]), and have shown that it can improve both communication and coordination. Similarly, a small body of work has emerged in recent years that explore eye-tracking technology as an input for digital games. Primarily this has concerned the creation of novel gaze-based gameplay experiences [39, 41], with only a few studies to date looking at shared eye tracking in competitive multiplayer games including our own [27, 34, 35, 40]. Typically, players of online games cannot see their opponents and so do not have access to the kinds of immediate behavioural cues (such as gaze) that enrich the experience of playing games in a physically shared location. In our previous work on the subject, we found that adding a visualisation of an unseen player's gaze into a digital game allowed spectators to make accurate predictions of what the player was planning to do, based on the areas of the screen receiving the most visual attention [34].

However, digital gaze visualisation is not a perfect replacement for face-to-face gaze observation. Researchers have found that live gaze visualisation from an eye tracker can be distracting and difficult for the observer to interpret, especially when there is a mismatch between gaze and other forms of communication [11, 30]. Furthermore, HCI researchers have found that the intended meaning of gaze behaviour can be difficult to discern through a visualisation [23], as it is often "hard to distinguish between deliberate communication and unwanted inputs" [33].

Gaze and Deception

As we use gaze cues to predict the intentions of others, gaze can be used against us as a method of deception. Players in many games mislead their opponents by simulating a fake intention with their gaze, as in a footballer's "no look pass". In *LaserViz*, the researchers simulated this mechanic using a two player shell game variation using cards, with one player wearing a head-mounted laser (that corresponds to the wearer's gaze) acting as the mixer [40]. The wearer initially experienced a disadvantage as the opponent could correctly guess their plans based on the gaze visualisation but in subsequent rounds, the advantage was reversed as the wearer learned to trick their opponent by consciously looking at the wrong card.

The effectiveness of misleading gaze behaviour is predicted by Levine's *Truth-Default Theory* [29], which holds that humans have a strong tendency to presume that another person is communicating honestly, and that lies are usually only detected after-the-fact based on contradictory evidence rather than through observation of the liar's demeanour. Levine explains this as an adaptive tendency for most situations, as lies are much rarer than truths, but one that leaves humans vulnerable to deceit. We believe this theory can be generalised to scenarios in which the communication is solely nonverbal.

Studies have found that humans have a limited ability to differentiate between true and deceptive gaze behaviours (e.g. [16]) despite developing this ability at young age. Freire et al. found that children as young as four years old could detect small "leaked" gaze cues, and use them to infer that a person's private thoughts were contradictory to their verbal statements [17]. Then again, people are also able to adopt honest-seeming gaze behaviour when they lie to avoid suspicion. For example, Mann et al. found that people maintained eye contact when lying to foster trust, by taking advantage of the popular belief that liars avoid eye contact [32]. This suggests that distinguishing between true and deceptive gaze signals is a difficult task that could provide a rich resource for challenging social play.

Gaze and Plan Recognition

The problem of plan recognition has long been investigated by artificial intelligence researchers. Kautz and Allen define plan recognition as "the task of inferring intentions (in terms of plans) from the actions or utterances of observed actors" [25]. Plan recognition is an important aspect in real-time strategy games, as human players who discern the plans of their opponents can gain a significant advantage. Further, players with knowledge of their opponents plans will allow them to employ some form of deceptive behaviour. Cohen et al. outline two types of plan recognition, *keyhole* and *intended* [9]. *Keyhole recognition* is the inference of an agent's goals and plans through unobtrusive observation, as if through a keyhole. The agent has no intent to communicate its goals to the observer. *Intended recognition* is the conveyance of an intentionally constructed impression of an agent's goals to the observer; it is typically found in adversarial settings, such as warfare, in which deceptive strategies are employed to thwart the recognition of the agent's true plans [8].

Strategic games have commonly been used as a context for adversarial plan recognition research (e.g. *StarCraft* [24]). Likewise, they have also been used in automated deception detection studies (e.g. *Mafia* [12]), which is a challenging problem that requires the combination multimodal inputs to effectively detect [1]. Although a large body of literature exists in these research areas, our interest for the purposes of this paper lies in plan recognition by human subjects. We seek to test whether a player who follows a live gaze visualisation will obtain the same benefits as *gaze following*, enabling them to discern their opponent's intentions and plans better.As previously mentioned, players in online games lack access to social cues, limiting their ability to discern plans until sufficient explicit information is provided. Byom and Multu explain that are three mechanisms of *Theory of Mind*: (1) knowledge of shared contexts, (2) perception of social cues, and (3) interpretation of actions, all of which are crucial when attempting to discern future behaviours [6]. The use of an online strategic game with incomplete information combined with the introduction of gaze visualisation allows us to study this effect. Furthermore, such games typically require players to construct hypotheses plans (that we can measure), and at the same time invite treachery, trickery and deception [37]. Additionally, playing strategic games have been found to activate regions of the brain associated with *Theory of Mind* [18]. Our work lies at several crossroads, as we begin to explore the potential of gaze as a playful nonverbal avenue for deception while drawing on its roots in face-to-face human communication.

AIMS

We aim to explore the uses of gaze visualisation in online multiplayer games by understanding its potential usage for intention-reading and deception. Gaze, being a key signal for *Theory of Mind*, can assist a player to accurately imagine the mind of their opponent and thus help them to plant false beliefs. We formed the following hypotheses about this activity, based on our review of the literature:

[H1] Players who have knowledge of their opponent's gaze are better at discerning their opponent's overall plans.

[H2] Players who are aware that their gaze is being watched will alter their gaze behaviour to deceive their opponent.

[H3] When a player knows their gaze is being watched and attempts to mislead their opponent through deceptive gaze behaviour, the opponent will not detect the deception.

[H4] Players will be less accurate at predicting the plans of their opponent when the opponent knows their gaze is being watched than when the opponent does not know their gaze is being watched.

We aim to build upon our previous work [34] in three ways:

(1) By using players instead of spectators, where players will make use of gaze visualisation of their opponent during live gameplay, balancing the tasks of devising their own strategic plans with making inferences about their opponent.

(2) By assessing the ability of players to interpret and respond to opponents' gaze cues over the course of a whole game, with a particular interest in the ability to make long-term predictions during the early game before in-game actions have been taken. Previously, we were only able to obtain short-term predictions from the short clips used.

(3) By introducing a deception aspect where one player is aware that their gaze is being visualised, to create an opportunity for the player to deliberately manipulate their gaze behaviour to potentially mislead their opponents.

STUDY

We selected the computer game *Ticket to Ride* following the success of its use in our prior work [34]. In this game, players compete to build train routes between adjacent cities across a map of North America, and gain points for building a connection between specific pairs of cities that they are assigned on "ticket" cards (goal cities e.g. *Vancouver - Montreal*), which are unknown to the opponent. Players lose points if they fail to complete ticket cards they hold, and can block each other's paths, as only one player can claim each train route. Therefore, players must plan their train networks carefully to minimise the risk that an opponent will block them by claiming important routes first. More detailed information on the rules of *Ticket to Ride* can be found on the game's website[1].

There are several factors that make *Ticket to Ride* a suitable game for this study, in addition to keeping continuity with our previous study [34]. As a turn-based game, it gives players time to observe their opponent without having to focus on their own actions concurrently. It contains hidden information in the form of secret goals. As a map-based game, it encourages players to look at locations that correspond with their plans—and provides opportunities to do so misleadingly. Finally, it has a short gameplay time (around 30 minutes or less for two players), and is sufficiently simple that players are able to learn the rules and develop an understanding of the general strategies within a few minutes. We structured our study to ensure both players were familiar enough with the game to make predictions about their opponent, and that Player A (P_A) would be familiar with the gaze visualisation format by the third round (see Figure 1) to allow them to develop ideas about how to use it against their CU-GA opponent.

[1] `http://www.daysofwonder.com/tickettoride/en/usa/`

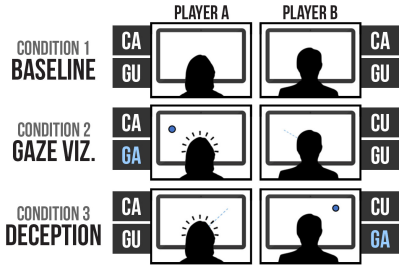


Figure 1: Study Conditions. In C1, no gaze data is visible to either player. In C2, P_A can see P_B's gaze, and P_B is unaware of this. In C3, P_B can see P_A's gaze; P_A is aware of this, and that P_B does not know P_A is aware.

Study Conditions

Each pair of players played three rounds of *Ticket to Ride* against each other, with each player being subjected to a condition (see Figure 1). Each condition consists of the players' awareness of the study condition (**CA**: Condition Aware, **CU**: Condition Unaware) and their awareness of their opponent's gaze (**GA**: Gaze Aware, **GU**: Gaze Unaware). Condition-Aware players were given all details of the study condition for that round, including what their opponent knew. Each players' awareness differed between rounds, as described below. One player was aware of the condition in every round (P_A), while the other was naive after the first round (P_B). The second and third round correspond to *keyhole* and *intended* plan recognition scenarios, respectively [9].

Condition 1 (Baseline)–No Gaze Visualisation of Players

The first round consisted of a standard online match with no gaze visualisation and both players are aware of this, making both players CA-GU (see Figure 1). This round served as a baseline for recording the players' behaviour under normal gameplay conditions. It also allowed players to get experience playing the game to minimise any learning effects, and to get a sense of forming their own strategy while discerning their opponent's strategy, based only on explicit in-game actions.

Condition 2 (Gaze Viz.)–Gaze Visualisation of Unaware Player

In the second round, P_A was shown the gaze of P_B throughout the game, and knew that P_B was not aware their gaze could be seen (see Figure 1). This allowed P_A, as a CA-GA player, to try to discern the intentions of an entirely naive opponent (CU-GU) through their gaze natural gaze behaviour.

Condition 3 (Deception)–Gaze Visualisation of Aware Player

In the third round, P_B received the gaze visualisation of P_A. But this time P_A knew their gaze was being tracked (see Fig-

ure 1). Not knowing this, P_B (CU-GA) attempted to discern the intentions of P_A (CA-GU) through their gaze behaviour without knowing that they could be intentionally misled.

Recruitment

We recruited 40 players (22 female) from the University of Melbourne, and assigned them into 20 pairs. These included undergraduate students, graduate students and non-faculty staff, aged between 18 and 39 years (M=23.2). Players were compensated with a $20 (AUD) gift card for their time. Before the study, players were asked to learn the rules of the game by watching a video tutorial, if necessary; 11 had played the game previously. A majority self-identified as an "occasional gamer" (23), while the rest self-identified as a "non-gamer" (11) or "regular gamer" (6). Ten players had used eye tracking previously, all as participants in a research setting.

Experimental Setup

The study was conducted in a university usability lab with two players in each session under approval by the University's ethics committee. To simulate online gameplay, we placed players in separate observation rooms and set up an online game of *Ticket to Ride* between them. We provided snacks and bottled water during the study in addition to the compensation due to the duration of each session. Both computers were fitted with a Tobii 4C eye tracker[2], and the gaze data from the eye tracker was streamed between the computers using a custom networked system that overlaid real-time gaze visualisations over any application (also used in our previous study [34]). Players were also given a tablet computer that mirrored their desktop computer screen and instructed to sketch notes about their opponent's strategy on the tablet using a stylus. We placed a video camera in each room to record the players, and the computer screens were also recorded for the duration of each session. Lastly, we logged the raw gaze data of both players for each game round for later analysis.

Procedure

Two researchers were present in each session to facilitate the study and to observe the players. The researchers followed a well-rehearsed script, so as not to reveal the study conditions to the players. On commencement, both players were given an initial briefing together that explained that their gaze would be tracked throughout the study for the sole purposes of later analysis. Players were then randomly allocated into one of the two observation rooms, with one researcher in each. There, players were given a written overview of the study, consent form and basic demographic questionnaire to fill out. Following this, players were calibrated using default calibration procedure of the eye tracking device, and instructed to play the game's interactive tutorial until both players were satisfied that they understood the game for up to 10 minutes.

Players then played three rounds of *Ticket to Ride* against each other through an online connection. To ensure timely completion, each player was given a 15 minute cumulative time allowance for their total turns in each round; if either player ran out of time, we manually calculated the scores for that round. At the start of each round, we requested each player to pick all three randomly assigned 'ticket' cards 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. The researchers prompted the player occasionally on these topics, and took notes throughout the session on player comments or behaviours related to the study objectives. Researchers were permitted to respond to any questions regarding the rules of the game at any time during the session. In the conditions where the player could see gaze, we provided a short tutorial on how we would like them to annotate on the tablet—*circling* the cities they thought were their opponent's goals and *highlighting* the routes they thought their opponent planned to take (see Figure 2). The gaze was visualised as a dynamic real-time heatmap, as our previous study showed that this was the best representation for discerning the goals of a player based solely on a gaze visualisation [34]. The dynamic heatmap visualisation was semi-transparent, and the application allowed the player to click "through" it as normal.

At the end of each round, players completed a brief section of the questionnaire (see *Measures and Analysis* section below for questionnaire details) about the round they just played. At the end of all three rounds, both players were invited to add any final thoughts, and the CU-GA player was asked whether they suspected any deception in the study. Finally, both players were fully debriefed about the study conditions.

Measures and Analysis

Player Self-Assessment and Experience

We created two sets of questionnaires for players to complete following each round of the game, with slightly different questions for each. In both sets, the first three questions following all conditions asked the player to rate their ability to predict their opponent's short-term and long-term plans, and what aspects of the game they had found helpful in predicting their opponent's plans. These repeated measures allowed an assessment of the subjective effects of gaze visualisation and deception on intention recognition ability. We expected that players would make a higher number of predictions when gaze was visualised and that these predictions would be more accurate when the opponent was unaware their gaze was being visualised. To test this effect, we ran a multiple regression model on the questionnaire responses, with independent variables being gaze awareness (whether the opponent's gaze was visualised on their screen or not) and condition awareness (whether the opponent was aware that their gaze was being visualised or not).

For the second and third round, the questionnaire asked additional questions to assess the impacts of gaze visualisation. Both players were asked to rate their experience with the gaze visualisation following the round in which they saw it. After the third round, the CA-GU player was asked whether they attempted to deceive their opponent and how. To test whether the deception of the study had been maintained, the CU player was asked after the second round in what ways the game had been different to the first round; and after the third round, the penultimate question asked whether they felt they had been deceived about the study at any point.

[2] `http://tobiigaming.com/eye-tracker-4c/`

Prediction Performance

We are primarily interested in players' actual success at discerning the plans of their opponents, as opposed to their self-assessed abilities. To determine this, we first extracted the ground truth (all routes each player claimed) for each individual game for each player (120 games total). We then analysed the study recordings to list all instances in which a player made an inference about their opponent's plan, including both verbal statements and tablet annotations. We then categorised these into *goal predictions* (goal cities the opponent is ultimately connecting to, i.e. their secret objectives) and *subgoal predictions* (what cities or routes they intend to connect through to reach their goal cities). We measured each of these predictions against the opponent's ground truth, defined as the secret objective cities they were assigned by the game, the routes they ultimately claimed, and any routes they verbally declared they were considering but did not end up claiming. We included the latter because players could change their strategy in response to events during the game, particularly after being blocked by an opponent who had correctly guessed their plan. We excluded from the ground truth some incidental routes, primarily any last-minute building decisions in the final turn that were clearly not part of their planning before that point in the game.

We assessed players' prediction performance using three metrics: (1) their accuracy in discerning goals, (2) the extent to which they discerned their opponent's plans, and (3) the *distance* of each guess from the ground truth, in terms of both goals and subgoals. We evaluated players' goal and plan prediction performance by employing the $F_1 - Score$ ($F_1 = 2 \times (precision \times recall)/(precision + recall)$) metric from information retrieval literature [36]. We compared the number of goal cities that were guessed correctly against the total number of cities guessed (*Precision*), and against the number of cities that were in the ground truth (*Recall*). We then performed the same comparisons, replacing the number of correctly-guessed goal cities with the number of goal cities that had been guessed as subgoals. A $F_1 - Score$ value was calculated from both sets of comparisons. Lastly, we employed the *distance* metric from graph theory [42]. As the playing area of *Ticket to Ride* is laid out as a graph diagram, it is possible to determine an error distance by treating all cities as nodes in a graph, and measuring how many nodes the guess was away from the ground truth [31]. To simplify our analysis, we take the shortest distance between the guess to any goal (or subgoal) from the ground truth. Predictions that were adjacent to a goal (or subgoal) thus have a distance of 1, and predictions that correctly matched the ground truth have a distance of 0.

RESULTS

Each session lasted for at least two hours, and consisted of three rounds of *Ticket to Ride* between two players, for a total of 60 rounds or 120 rounds if each player is considered separately. We then compared the extracted predictions made for each player in all conditions against the ground truth for that respective game. The use of the "think-aloud" protocol in our study provided rich descriptions of players' strategies and considerations throughout each of the games. These reflections gave us confidence in our analyses of the game data, as the results are in accordance with what players said themselves.

Effect of Gaze Awareness and Deception

Effect on Player's Self-Perceived Predictive Ability

We examine players' self-perceived predictive ability, i.e. how well they *thought* they discerned the plans of their opponent. We built a multiple regression model to predict player's self-ratings on this measure. A significant regression equation was found for both short-term (R^2*=0.23, F(3,116)=11.28, p<0.05*) and long-term (R^2*=0.20, F(3,116)=9.42, p<0.05*) predictive ability. We can define the model based on the latter as follows:

$$Response = 3.27 + (1.8\,(Gaze)) + (-1.32\,(Deception))$$

Both measures are significantly correlated across all conditions for both players (*Pearson(r)=0.78, p<0.05*), meaning that if a player rated their ability to predict specific plans highly, their ability to predict overall goals was also rated highly. The model also shows that there were no learning effects in predictive ability, but there was an increase in players' confidence in their ability to form strategies and make predictions.

Effect on Player's Outcomes and Experience

We first report on players' self-assessments (questionnaire responses), each measured on a 7-point Likert scale (1 being lowest agreement, 7 being highest agreement). Subsequently, we report on the way players spoke about the experience of the game, and the level of success that they achieved.

Overall, players found it easier to make short-term predictions (moves) (P_A*=5.6,*P_B*=5.4*) and long-term predictions (strategy) (P_A*=5.1,*P_B*=5.3*) in conditions where gaze was visualised. Both sets of players noted the gaze visualisation was useful for formulating their own strategy (P_A*=4.9,*P_B*=5.1*), and that it changed the way they played the game (P_A*=5.4,*P_B*=5.2*). Both sets of players also believed that gaze visualisation had an influence on the outcome of the game (P_A*=5.1,*P_B*=4.7*). In Condition 3, the CA-GU players indicated that they changed the way they played the game as they knew that their opponent could see where they were looking (*M=5.2,SD=1.6*). Consequently, the players were somewhat bothered by the knowledge that their gaze was being observed (*M=3.9,SD=1.7*), and all of them employed some form of deceptive strategy to counteract this (see *Gaze-Based Deception Strategies* section).

In a comparison of the actual game outcomes in Conditions 2 and 3, we found that the CA players were no more successful than the CU players. This was despite their information advantage, as demonstrated by their greater ability to discern the opponent's true intentions (see Table 1). To explain this, we refer to the comments from the CA player for both conditions. In Condition 2, CA-GA players commented that the gaze visualisation was distracting and drew their attention away from their own plans, both by providing additional information about their opponent's plans and by creating visual clutter on the screen. We believe that this was likely exacerbated by the fact that real-time gaze visualisations were an unfamiliar interface element for all players. The CA-GA players also commented that the visualisation influenced them towards a 'blocking' strategy, which was focused on interfering with their opponent rather than pursuing their own goals. This strategy proved relatively ineffective in many cases, as it caused the aggressor to waste as many turns and resources as the player they were trying to block. For example, $P6_A$ stopped their opponent

				Goal Prediction		Plan Prediction		Distance	
Condition	*Player*	*Won*	*Avg. Predictions*	$F_1 - Score^1$	$F_1 - Score^2$	*Precision (Recall)*	$F_1 - Score$	*Total*	*Mean*
Cond. 1 (Baseline)	P_A (CA-GU)	13	5.8	0.06	0.13	0.49 (0.13)	0.19	25	0.22
	P_B (CA-GU)	7	4.5	0.11	0.19	0.51 (0.11)	0.17	29	0.32
Cond. 2 (Gaze Viz.)	P_A (CA-GA)	11	12.7	0.23	0.44	0.48 (0.34)	0.39	75	0.30
	P_B (CU-GU)	9	5.8	0.02	0.04	0.49 (0.13)	0.18	24	0.21
Cond. 3 (Deception)	P_A (CA-GU)	12	4.0	0.09	0.17	0.48 (0.09)	0.14	24	0.30
	P_B (CU-GA)	8	11.7	0.18	0.35	0.41 (0.28)	0.31	111	0.48

Table 1: Prediction Performance Results. The F_1–$Score^2$ is calculated based on all cities (goals and subgoals) guessed against the ground truth to find out if players did discern the goal cities did not recognise them as goal cities while F_1–$Score^1$ is calculated based on only correctly guessed goal cities.

from completing 2 out of 3 ticket cards, but completed none of their own objectives. The CA-GU players spoke about the difficulty of maintaining their deceptive strategies while also thinking about their plans, especially early in the game, noting that it was the best time to be deceptive as their opponent had no other cues. This difficulty in maintaining deception echoes the findings of the *LaserViz* study [40]. Many of the CA-GA players commented that they felt like they were 'cheating' in the second round with an 'unfair advantage', whereas the final round felt more challenging, and generally more exciting. Nevertheless, the CA-GU players found it interesting and engaging to know that their gaze was being watched, and stated that it gave the game added depth.

At the conclusion of each study session, the deceived player was asked whether they suspected that they had been deceived at any point. Only one player expressed some suspicion while the rest did not express any at all. All players expressed the opinion that both they and their opponent had improved their performance each round, which they attributed to ordinary factors such as the luck of the draw or taking greater care in planning their strategies.

Effect on Prediction Performance

Table 1 displays a summary of prediction performance results for all three conditions. A total of 889 predictions (guesses) were made by both players across all conditions. The introduction of gaze increased the average number of raw predictions made compared to when players could not see gaze. Further, we noted that players were able to make their first predictions significantly earlier when given the ability to visualise the gaze of their opponent than in the other conditions without gaze, on average within the first minute from the start of the game. In the conditions without gaze visualisation, the first prediction on the opponent's plans was only made after the opponent claimed a few routes, typically a few minutes into the game. Predictions did become increasingly difficult if an opponent was effective in hiding their intentions. On occasion, the researchers would prompt the player on what they thought their opponent's plans might be, which would often result in a negative response from the player (e.g. *"I don't know", "I am not certain", etc.*), or a brief response with no solid prediction that the researcher could record.

A reason for low prediction counts, particularly in the conditions where players were gaze unaware, is that players often focused on their own plans initially and were less concerned about their opponent's plans until completing their own. This was more likely in round 1 as new players got used to the game. Players may have been particularly good at hiding their plans, in line with the game design, making it hard to make predictions. This led to the predicting players only making predictions when they were confident, leading to a low number of average predictions with high precision but low recall.

In the conditions where players were gaze aware, they explicitly mentioned that they used the gaze visualisation extensively to make predictions, especially in the early stage of the game. This was as expected, as few actions are played in the early stage and the formation of their own plans was still in its infancy. We highlight that players used a variety of in-game actions as well to make predictions, such as observing which colour cards the opponent picked from the deck and matching that to the colour of routes on the board, or observing the general direction in which the opponent was building. Experienced players used the average distance between pairs of cities to gauge whether their opponent had completed a ticket. Other in-game cues contributed to their predictive ability, such as observing which train cards were not selected and analysing opponent's building patterns. Players who frequently annotated on the tablet found it effective for keeping track of their opponent's plans.

In the third condition, the CA-GU players made fewer predictions about their opponent's plans than in the previous rounds. We believe this is for three reasons. Firstly, players focused first on deceiving their opponent and second on attempting to complete their own goals, leaving little attention for discerning their opponent's plans (*"I wasn't focused on predicting their movements but rather on how to throw them off since I know I'm being watched."* [$P9_A$]). Secondly, the deceptive gaze activity required these players to keep track of two sets of plans (true and deceptive), making it hard to keep track of a third. Thirdly, players felt less confident about their predictive abilities after seeing the gaze visualisation and then having it removed (*"I was less able to assess my opponent's strategy than I thought I could. It allowed to see the real advantage of having the eye tracking in the second round. With the eye tracking, I would have been able to see if she was looking at the east coast the entire time."* [$P11_A$]).

Upon calculating the different metrics to determine the effect of gaze and deception on prediction performance, we found that the CA-GA players performed the best overall in Condition 2 amongst the other rounds, despite having the highest number of predictions. There was some degree of error indi-

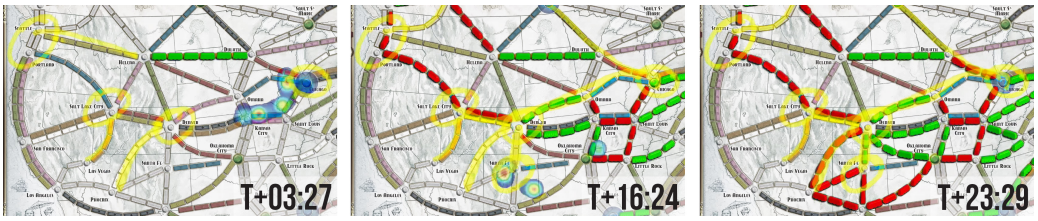


Figure 2: Ticket to Ride Gaze Visualisation Prediction Example. The three figure above shows three stages (early, mid, late) of the same game from left to right, and the real-time heatmap visualisation used in our study. The green tracks belong to the CA-GA player (who has annotations in yellow highlight) while the red belongs to CU-GU player. In early stage, the CA-GA has already predicted part of their opponent's plans, discerning that the opponent is interested in a particular route(s) when no trains have been played yet by the opponent. In the mid stage, we can see how the predicted strategy has been played out and now better discern a different part of their plan. In the late stage, we can see more of the predictions materialising. The figure has been enhanced by reducing the contrast of the background to better show the real-time visualisation, paths claimed and the annotations.

cated by the total distance (75), but this appears to be in the normal range when averaged, meaning that the predictions made did indeed have a distance but was of low value. In contrast, performance scores for the CU-GA players in Condition 3 reveal that they still made more correct predictions based on the goal and plan prediction scores compared to conditions with no gaze. Further, the distance metric did reveal that only 36.6% of the CU-GA players predictions had a distance of over 1 (CA-GA was 26%), showing that the players could still judge their opponents' plans with reasonable accuracy.

We expected the likelihood of players discerning the goals of their opponents to be generally low, especially in conditions without gaze, and this proved to be true ($F_1 - Score^1$). As a further step, we wanted to see if players would guess their opponents' goals as part of a route (i.e. as subgoals) without discerning them as end goals (i.e. without circling them or stating that they were a 'ticket' destination). The results for this analysis were higher ($F_1 - Score^2$), meaning that players did indeed discern goal cities without recognising them as end goals. A further comparison revealed that CU-GU players in Condition 2 had a constant low prediction score, which we believe is attributable to three factors. First, the CA-GA players kept their plans hidden better, by playing a more reserved strategy. Second, the CA-GA players often employed a blocking strategy, causing the CU-GU players to make incorrect or fewer predictions. Third, the CA-GA players completed less of their own intended plans in this round, so there was less on-board information for the CU-GA players to make inferences from. As an additional point, we observed that players commonly extended their routes beyond their goal cities to get bonus points (*Ticket to Ride* gives bonus points to the player with the longest overall route), which made it harder to discern whether the 'ticket' goal was the end of a route or not. In Condition 3, CA-GU players (e.g. $P15_A$) commonly looked beyond their goal cities as a form of *gaze misdirection*.

The CA-GA players in Condition 2 scored an average plan prediction score of 0.39 (*SE=0.031*) when they could see their opponent's gaze, compared to 0.19 (*SE=0.036*) in Condition 1 when they could not. A Welch Two Sample t-test indicated a significant difference between rounds (*t=4.20, df=37.23, p<0.05*), along with a medium-sized effect *r=0.57*. We thus reject the null hypothesis [H1]. The CU-GA players in Condition 3 scored an average distance error of 0.48 (*SE=0.050*), whereas the CA-GU players in Condition 2 scored an average distance error of 0.3 (*SE=0.033*). A Welch Two Sample t-test indicated a significant difference between rounds (*t=3.01, df=413.32, p<0.05*), with a small-sized effect *r=0.15*. We thus reject the null hypothesis [H4].

Summary

Our results showed that gaze added value when used to discern plans in a strategic game, especially early in the game, consequently revealing a large part of the tracked player's plans as the game progressed. Figure 2 illustrates how gaze was displayed, followed by its early and partial prediction of an opponent's plan. Further, we found that deception did have an effect on predictive performance, though the player's true intentions still leaked through their gaze. The CA-GU players did need to look at their real plans on occasion and, over time, gave them away as they claimed more routes. As a result, CU-GA players guessed both true and deceptive plans. The results are also consistent with Hypothesis [H3], as the CU-GA players did not state that they were being deceived (as predicted by the *Truth-Default Theory* [29]), when in fact all the CA-GU players did attempt to deceive. Gaze visualisation had overall positive effects on both player experience and player performance, although it was also found to be 'distracting'. The findings also show that the deception condition did cause the CA-GU players to change their gaze behaviour, as they became more aware of where they were looking and sought to deceive their opponent, validating Hypothesis [H2]. The strategies employed are summarised in the section below.

GAZE-BASED DECEPTION STRATEGIES

All CA-GU players in the final round stated that they attempted to deceive their opponent by manipulating their gaze behaviour, and described one or more strategies they used to do so. We categorised the strategies according to the general theory of deception, which divides deception methods into two main categories: *dissimulation* (covert, hiding what is real) and *simulation* (overt, showing the false) [4, 5]. Within these categories, we have created a more specific typology that draws on both the general theory and deception concepts in military strategy, such as camouflaging [13].

Dissimulation Deception Strategies

We observed several ways that players sought to conceal their plans when they knew their gaze was being observed. The most frequently employed and direct dissimulation strategies

were *Gaze Averting* (6) and *Information Reduction* (6), followed by *Gaze Scattering* (5). The least employed strategies, *Risk reduction* (3) and *Obstruction* (4) were more oblique, seeking to deny the opponent an accurate target to uncover.

Gaze Averting: Players sought to avoid looking at the routes and cities that were of interest to them. In Bell and Whaley's typology, this was a kind of *masking* [5]. The practice was difficult to sustain; players frequently expressed frustration at the conflict between their desire to see what they were doing and their anticipation that doing so would give away their plans to the opponent. From observing and listening to players, it was apparent that this avoidance tactic took concentrated effort to maintain, and as such, it tended to lapse for increasing lengths of time as the game went on. In face-to-face communication, gaze averting is often attributed as a tell-tale sign of someone lying [14], but this was not detected in our study.

Gaze Scattering: Players went a step further than aversion, by actively scattering their gaze around the screen to give the impression that they were looking "everywhere". This tactic was intended to make it hard to distinguish moments of real planning from moments of arbitrary looking—similar to Bell and Whaley's concept of *dazzling* [5]. Gaze scattering also served as a kind of attack on the opponent's concentration. In Condition 2, all of the CA-GA players noted that the gaze visualisation distracted them from planning their own strategy under time pressure. In turn, players replicate this distraction for their opponent in Condition 3 by scattering their gaze, essentially "weaponising" the gaze visualisation.

Information Reduction: Players employed several tactics to reduce the amount of information their actions could give away. For example, players favoured drawing cards from the blind deck over drawing face-up cards, so that the opponent could not see what colours they were collecting. Other players took their turns as quickly as possible, minimising their opponent's time-frame for making inferences about their gaze and actions, causing the opponent to concentrate on their own plans instead.

Risk Reduction: Players adopted a more flexible strategy when they were aware that their gaze was being observed, to minimise the effect of being blocked by their opponent. For example, players chose to build across areas that offered the most alternative pathways, rather than going for a more direct and efficient route. Players who adopted this strategy typically tried not to claim any train routes for the first half of the game, and focused instead on collecting cards; this both reduced the information available to the opponent (gaze and in-game), and allowed the player to re-plan their route if the opponent's actions got in the way of their initial plan. Risk reduction strategies such as these ensured the player's intentions could not be pinned down, because they were left open to change.

Obstruction: Players sought to disrupt their opponent as a form of counter-attack, taking the focus away from their plans by playing aggressive moves such as taking cards and routes that they thought their opponent might need. The intended benefits of this were twofold. First, if done well, it forces the opponent to use up their resources to work around the blocking moves, which ideally left the opponent unable to complete their objectives by the game's end. Second, by forcing the opponent to play reactively, it neutralises the opponent's information advantage as the opponent would be less concerned about the gaze-tracked player's secret objectives. This was a high-risk strategy, as the player attempting it had less information available to them to know whether their aggressive moves were, in fact, obstructing their opponent's plans.

Simulation Deception Strategies

As well as hiding their true intentions, players sought to actively mislead their opponent through various kinds of *Gaze Misdirection* (3), a specific example of which was *Decoys* (16). Players also used *Gaze Camouflage* (4) strategies to create a false impression about the meaning of their actions.

Gaze Misdirection: Players attempted to give their opponent a false idea about what they were thinking by exhibiting behaviour that was consistent with an imagined alternative plan. For example, the player might stare at a route they think the opponent wants to build on, to scare them into thinking they will be blocked; this was intended to induce a degree of panic and to keep them from noticing the player's real plans. At times, players would take this further by alternating fixations between the route and a face-up card of the matching colour, to heighten the impression that the player was planning out their own construction, rather than idly looking at the map.

Decoys: A common sub-case of gaze misdirection was the use of decoy paths and is the most common deception strategy employed. Players who employed this strategy chose a set of cities far away from their real objectives, and repeatedly traced a path between them to give the impression that they were planning to build there. The decoy was usually chosen at the beginning of the game, shortly after the player had received their real objectives. Players typically watched out for their opponent building on their decoy path, and took this as confirmation that they had deceived the opponent. For the disinformation to be effective, however, the enemy must know how it is being accepted and interpreted by the target [22].

This strategy appears to be the most effective, particularly as opponents showed a degree of confirmation bias in their observations. Typically, the opponent expressed an early suspicion about one or two locations that the CA-GU player had looked at, and subsequently, they would notice any time the player's gaze overlapped that point, while being less alert to other areas that the player was looking at. This meant that opponents who were deceived by an early decoy often remained distracted by it, even after the CA-GU player has moved on from the decoy. Below is a chronological example from Session 17 ($P17_B$):

[03:09] *"She's really focusing on Pittsburgh."*

[05:57] *"I'm still convinced she's going to go Pittsburgh somehow. She won't stop looking at Pittsburgh."*

[08:39] *"She wants to do a loop down underneath. I just don't know how she plans to get to Pittsburgh after that."*

[09:02] *"Ok now she wants to get to Pittsburgh from here" [after opponent claimed Helena to Denver route]*

[11:49] *"I still think she is trying to get to Pittsburgh."*

[12:46] *"Maybe she didn't need to go Pittsburgh at all." [after opponent started to take new ticket cards]*

Gaze misdirections and decoys became less effective as the board filled up with claimed routes, as these were taken as more reliable evidence of the players' intentions than the gaze tracking. To counter this, there are instances where CA-GU players even spent resources to claim a route along their decoy path, just to sell the idea that they were planning to build there.

Gaze Camouflage: Players used elements of the game board as cover for their real point of attention. This corresponds to *repackaging* in Bell and Whaley's typology [5]. This strategy involved the player fixating for an extended period on routes that had already been claimed, to give the impression that they were considering this existing network, when in fact they were using their peripheral vision or brief side glances to plan a new path on nearby unclaimed routes. At other times this was done with elements outside the main game board, such as the cards around the outside. In some cases, the player appeared to be unaware that their face was turned towards the intended "cover point" but their eyes were still directed at their real goal, so the deception was unsuccessful.

DISCUSSION

Players demonstrated a deep understanding of how to utilise gaze visualisation and their added knowledge about their opponent's plan. This went beyond simply discerning the opponent's intended paths and blocking them. A common tactic was to plan the timing of an action based on its proximity to the opponent's gaze: if the player saw the opponent looking in the vicinity of a path they wanted, that path became the player's first objective; other paths were de-prioritised relative to their distance from the opponent's area of focus. If the opponent's gaze was far away from all of the player's goals, the player was more likely to build up resources and bide their time before claiming any routes, allowing them to play in an efficient manner that avoided signalling their own plans to the opponent. When players were confident in their predictions, they were able to plan for deception. For example, $P9_A$ predicted a multi-step path that their opponent would take, waited until the opponent had built along most of that path (confirming the prediction), and then claimed the final section at the last moment, forcing the opponent to take a long detour after they had invested many resources into the original route. Gaze unaware players were much less likely to engage in such aggressive tactics, as they were less confident that their actions would be in the to stop or disrupt their opponent.

The benefits of the gaze visualisation were offset by a cost in players' time and attention. Players reported that the visualisation was distracting, and made it hard to concentrate on planning their own strategy: *"If my task is only to predict my opponent's moves, that would be substantially an easier task, especially with the eye tracking. But when you try to make predictions and play the game with a time constraint, the cognitive load starts to build up."* [$P2_A$]. Accordingly, we observed that players who could see the gaze visualisation spoke less about their own strategy and more about their opponent. Players who could not see the visualisation were more likely to talk about the opponent's strategy in terms of potential risks and contingencies for their own strategy. Further, players speculated that the visualisation might be a net disadvantage in terms of winning the game due to its distracting nature, although most took a guess and mentioned that it would become more advantageous once players became accustomed to it.

The level of distraction may be partly attributable to the visualisation style. A few players commented that they would prefer a simple moving dot to the dynamic heatmap used in the study. The choice of the dynamic heatmap was informed by our prior study [34], in which it was the most accurately interpreted and second-most preferred visualisation style among nine formats compared. However, participants in that study were observers of a game rather than active players. It is possible that a less information-rich visualisation would be more manageable for players, who need to focus on developing their own plans as well as discerning those of an opponent. Moreover, the prior study showed substantial variations in preferences between participants. All in all, this suggests that it would be beneficial to give players some control over the gaze visualisation style, such as the ability to change its appearance or switch it on and off to manage the flow of information.

Non-live gaze visualisations could also help to avoid the distraction effects observed in this study. Mechanisms can be implemented to give each player a summary of where their opponent in an online match has looked, and for how long. For example, in *GeoGazemarks*, fixations were recorded and clustered to give a history of a user's points of interest to support free map exploration [19]. This is similar in function to the dynamic heatmap visualisation used in this study, although our heatmap provided a persistent summary only of gaze behaviour over the past few seconds; this persistent-summarising quality was one of the characteristics for which it was favoured in our prior study [34]. However, a non-live gaze visualisation would lose some of the affordances that provided for interesting gameplay opportunities found in our study, such as the use of *gaze scattering* strategy to distract an opponent.

While our study supports the general consensus of prior literature that gaze visualisations can be disruptive as well as informative, it also points to an opportunity for game designers to utilise this as a double-sided resource for players. We provide three examples of how this resource might be used as a "power up" in online multiplayer games. First, we can provide live information about where an opponent is looking, at the cost of distracting visual noise and the risk of being misled by false signals. This can also take the form of non-live gaze visualisations, giving the player a summary of the opponent's gaze but leaving open the possibility that the opponent had manipulated their gaze to be deceptive. Second, we can limit where gaze visualisations appear; either activated by the player, activated arbitrarily (such as on a timer), or activated by other gameplay systems (for example, only appearing in areas that the player has previously explored, similar to *fog of war* in concept). Allowing the player to see gaze visualisations in only part of the game could leave them with a strategic choice of whether to seek further gaze information, at the potential cost of distraction, or allow themselves to be oblivious to their opponent's gaze. Third, we can use awareness of gaze sharing as a game resource, so that under certain conditions a player is notified when their opponent can see their gaze, allowing them to undertake deception strategies.

Generalising the implications of this study to other types of games should be done with caution, however. *Ticket to Ride* was chosen as an appropriate game for gaze sharing as it is turn-based and both players see a largely identical screen, so they can watch each other taking turns. These qualities are common to many digital board games, but are not present in many popular genres of online games. For example, in a real-time strategy game such as *Starcraft*, players scroll across a game map that is larger than the boundaries of their screen. A gaze visualisation for this kind of game would have to be anchored to the in-game location that a player is observing, rather than just the coordinates on the screen. This would reduce the amount of time each player spent observing their opponents' gaze, as it would allow them to monitor gaze activity only in areas they are concerned with, or only in areas they have previously explored (as in fog of war). A player may deduce, for example, that their opponent is not launching an attack as long as their gaze activity is isolated to the area of their own base. However, in a real-time game such as this, we assume the level of distraction caused by gaze visualisation would be even more acute than in a turn-based game as players do not have time to sit and think.

Players were successful in devising deception strategies that reduced their opponent's prediction accuracy about their plans, but also found degrees of difficulty in maintaining them (in line with [40]). All of the strategies we observed were conceived independently by the players, without suggestions from the researchers, and the stated inspiration behind many of the strategies was the player's own experience of observing the gaze visualisation. Players were largely successful in converting their own experience into an accurate enough model of their opponent's thinking that they could implant a false belief about their strategy, or disrupt the prediction process in other ways. This demonstrates the application of *Theory of Mind*, as the *ability to deceive* is considered one of its hallmarks [28].

The effects of gaze visualisation were most pronounced in the early stage of the game. This was true both for the ability to make predictions and for the ability to deceive the opponent. At this stage, gaze data is a relatively clear signal, as there is little other information available in the form of routes claimed on the board. As the game progresses, deception through altered gaze behaviour becomes more difficult, as the accumulation of claimed routes on the board provides stronger evidence of intention than gaze behaviour. Claimed routes act as a costly signal due to the investment of resources involved, and so where there is a mismatch between routes and gaze (lack of correspondence, in Levine's terms [29]), the gaze information is ignored in favour of the route information.

This study extends prior work (e.g [34, 40]) by demonstrating that human subjects can make long-term predictions and employ deception given the ability to visualise gaze for longer durations. However, players commented that they lacked the attention and memory capacity to fully utilise the information that was theoretically available to them through the gaze visualisation. From our human observations, we found that players in our study ignored large areas of the board (e.g. whole left side) when they could see that their opponent was not looking at it. Based on their comments, it is apparent that this ability to triage the board on a wide scale was a substantial part of the competitive benefit players derived from gaze visualisation, comparable at least to the more fine-grained prediction of specific routes. We believe that future gaze-aware systems can do the same, by ignoring elements that are irrelevant to minimise computational cost. Our study also showed gaze enabled players to make earlier and more accurate predictions, which have positive implications for automated plan recognition systems [38]. Researchers have demonstrated that systems that can predict intentions based on gaze patterns [21], but beyond the question of how well computer systems can predict human intentions through gaze, there is the question of whether these systems could detect deception in human gaze behaviour, and distinguish "true" intentions from false.

Limitations

Whereas our previous studies have looked at gaze interpretation by passive observers (spectators) [34], this study involved gaze interpretation by active players. This allowed players to forestall their opponents' plans, potentially changing the outcome. As a result, a degree of human judgement was necessary to determine whether predictions were accurate at the time they were made, as players sometimes had to deviate from their original plans due to events in the game. Players used a combination of the gaze visualisation and in-game actions for their predictions, and we did not discriminate between guesses that were largely informed by one or the other.

During the study, players were prompted for their predictions only when they had been silent for a period of time, rather than at regular intervals; this helped to preserve the naturalistic flow of gameplay, but meant that prediction intervals varied between players and rounds. Asking for predictions at fixed intervals would standardise the results and allow a more standardised comparison between conditions and players, at the cost of a somewhat more artificial gameplay experience. Conversely, our results are subject to the experimental design—prompting players to make inferences about their opponents' strategy is likely to have influenced how they reasoned about the game, to at least a small extent.

CONCLUSION

In a lab study of an online strategy game enhanced with one-way shared gaze visualisation, we found that players were able use gaze visualisation to improve their ability to predict an opponent's intentions. Players were also able to alter their own gaze behaviour deliberately to hide their true intentions and convey false impressions, to a degree. We defined a typology of gaze-based deception strategies, which reflects the general theory of deception [5]. However, this deceptive behaviour required sustained effort, and players could not prevent some of their intentions from 'leaking' to the opponent through their gaze. Players found this gaze-based deception task to be challenging and engaging, and we have identified several avenues that may be fruitful for game design to explore.

ACKNOWLEDGMENTS

We acknowledge the Australian Commonwealth Government and the Microsoft Research Centre for Social NUI for their support on this project. We would like to thank our colleagues, Yomna Abdelrahman and Vassilis Kostakos for their input.

REFERENCES

1. Mohamed Abouelenien, Veronica Pérez-Rosas, Rada Mihalcea, and Mihai Burzo. 2014. Deception Detection Using a Multimodal Approach. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, NY, USA, 58–65. DOI: http://dx.doi.org/10.1145/2663204.2663229
2. Deepak Akkil, Jobin Mathew James, Poika Isokoski, and Jari Kangas. 2016. GazeTorch: Enabling Gaze Awareness in Collaborative Physical Tasks. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 1151–1158. DOI: http://dx.doi.org/10.1145/2851581.2892459
3. Simon Baron-Cohen, Ruth Campbell, Annette Karmiloff-Smith, Julia Grant, and Jane Walker. 1995. Are children with autism blind to the mentalistic significance of the eyes? *British Journal of Developmental Psychology* 13, 4 (1995), 379–398. DOI: http://dx.doi.org/10.1111/j.2044-835X.1995.tb00687.x
4. Whaley Barton. 1982. Toward a general theory of deception. *Journal of Strategic Studies* 5, 1 (1982), 178–192. DOI: http://dx.doi.org/10.1080/01402398208437106
5. John Bowyer Bell and Barton Whaley. 1991. *Cheating and deception*. Transaction Publishers.
6. Lindsey Byom and Bilge Mutlu. 2013. Theory of mind: mechanisms, methods, and new directions. *Frontiers in Human Neuroscience* 7 (2013), 413. DOI: http://dx.doi.org/10.3389/fnhum.2013.00413
7. Andrew J Calder, Andrew D Lawrence, Jill Keane, Sophie K Scott, Adrian M Owen, Ingrid Christoffels, and Andrew W Young. 2002. Reading the mind from eye gaze. *Neuropsychologia* 40, 8 (2002), 1129 – 1138. DOI: http://dx.doi.org/10.1016/S0028-3932(02)00008-8
8. Sandra Carberry. 2001. Techniques for Plan Recognition. *User Modeling and User-Adapted Interaction* 11, 1 (01 Mar 2001), 31–48. DOI: http://dx.doi.org/10.1023/A:1011118925938
9. Philip R Cohen, C Raymond Perrault, and James F Allen. 1981. Beyond question answering. *Strategies for natural language processing* 245274 (1981).
10. Sarah D'Angelo and Andrew Begel. 2017. Improving Communication Between Pair Programmers Using Shared Gaze Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6245–6290. DOI: http://dx.doi.org/10.1145/3025453.3025573
11. Sarah D'Angelo and Darren Gergle. 2016. Gazed and Confused: Understanding and Designing Shared Gaze for Remote Collaboration. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2492–2496. DOI: http://dx.doi.org/10.1145/2858036.2858499
12. Sergey Demyanov, James Bailey, Kotagiri Ramamohanarao, and Christopher Leckie. 2015. Detection of Deception in the Mafia Party Game. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*. ACM, New York, NY, USA, 335–342. DOI: http://dx.doi.org/10.1145/2818346.2820745
13. James F. Dunnigan and Albert A. Nofi. 2001. *Victory and Deceit: Deception and Trickery at War*. Writers Club Press.
14. Shiri Einav and Bruce M Hood. 2008. Tell-tale eyes: children's attribution of gaze aversion as a lying cue. *Developmental Psychology* 44, 6 (2008), 1655. DOI: http://dx.doi.org/10.1037/a0013299
15. Paul Ekman and Wallace V. Friesen. 1969. Nonverbal Leakage and Clues to Deception. *Psychiatry* 32, 1 (1969), 88–106. DOI: http://dx.doi.org/10.1080/00332747.1969.11023575 PMID: 27785970.
16. Tom Foulsham and Maria Lock. 2015. How the Eyes Tell Lies: Social Gaze During a Preference Task. *Cognitive Science* 39, 7 (2015), 1704–1726. DOI: http://dx.doi.org/10.1111/cogs.12211
17. Alejo Freire, Michelle Eskritt, and Kang Lee. 2004. Are Eyes Windows to a Deceiver's Soul? Children's Use of Another's Eye Gaze Cues in a Deceptive Situation. *Developmental Psychology* 40, 6 (2004), 1093 – 1104.
18. Helen L. Gallagher and Christopher D. Frith. 2003. Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences* 7, 2 (2003), 77 – 83. DOI: http://dx.doi.org/10.1016/S1364-6613(02)00025-6
19. Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2012. GeoGazemarks: Providing Gaze History for the Orientation on Small Display Maps. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*. ACM, New York, NY, USA, 165–172. DOI: http://dx.doi.org/10.1145/2388676.2388711
20. Matthias S. Gobel, Heejung S. Kim, and Daniel C. Richardson. 2015. The dual function of social gaze. *Cognition* 136, Supplement C (2015), 359 – 364. DOI: http://dx.doi.org/10.1016/j.cognition.2014.11.040
21. Chien-Ming Huang, Sean Andrist, Allison SauppÃľ, and Bilge Mutlu. 2015. Using gaze patterns to predict task intent in collaboration. *Frontiers in Psychology* 6 (2015), 1049. DOI:http://dx.doi.org/10.3389/fpsyg.2015.01049
22. R Hyman and O Druckman. 1991. A broader concept of deception. *In the mind's eye: enhancing human performance* (1991).
23. Robert J. K. Jacob. 1990. 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 (CHI '90)*. ACM, New York, NY, USA, 11–18. DOI: http://dx.doi.org/10.1145/97243.97246

24. Froduald Kabanza, Philipe Bellefeuille, Francis Bisson, Abder Rezak Benaskeur, and Hengameh Irandoust. 2010. Opponent Behaviour Recognition for Real-time Strategy Games. In *Proceedings of the 5th AAAI Conference on Plan, Activity, and Intent Recognition (AAAIWS'10-05)*. AAAI Press, 29–36. http://dl.acm.org/citation.cfm?id=2908558.2908563

25. Henry A. Kautz and James F. Allen. 1986. Generalized Plan Recognition. In *Proceedings of the Fifth AAAI National Conference on Artificial Intelligence (AAAI'86)*. AAAI Press, 32–37. http://dl.acm.org/citation.cfm?id=2887770.2887776

26. Adam Kendon. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26 (1967), 22 – 63. DOI:http://dx.doi.org/10.1016/0001-6918(67)90005-4

27. Michael Lankes, Bernhard Maurer, and Barbara Stiglbauer. 2016. 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 (ACE2016)*. ACM, New York, NY, USA, Article 17, 9 pages. DOI: http://dx.doi.org/10.1145/3001773.3001774

28. Alan M Leslie. 1987. Pretense and representation: The origins of "theory of mind.". *Psychological review* 94, 4 (1987), 412.

29. Timothy R. Levine. 2014. Truth-Default Theory (TDT): A Theory of Human Deception and Deception Detection. *Journal of Language and Social Psychology* 33, 4 (2014), 378–392. DOI: http://dx.doi.org/10.1177/0261927X14535916

30. Jerry Li, Mia Manavalan, Sarah D'Angelo, and Darren Gergle. 2016. Designing Shared Gaze Awareness for Remote Collaboration. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, 325–328. DOI: http://dx.doi.org/10.1145/2818052.2869097

31. Darren Lim. 2007. Taking Students out for a Ride: Using a Board Game to Teach Graph Theory. In *Proceedings of the 38th SIGCSE Technical Symposium on Computer Science Education (SIGCSE '07)*. ACM, New York, NY, USA, 367–371. DOI: http://dx.doi.org/10.1145/1227310.1227439

32. Samantha Mann, Aldert Vrij, Sharon Leal, Pär Anders Granhag, Lara Warmelink, and Dave Forrester. 2012. Windows to the Soul? Deliberate Eye Contact as a Cue to Deceit. *Journal of Nonverbal Behavior* 36, 3 (2012), 205–215. DOI: http://dx.doi.org/10.1007/s10919-012-0132-y

33. Bernhard Maurer, Michael Lankes, Barbara Stiglbauer, and Manfred Tscheligi. 2016. *EyeCo: Effects of Shared Gaze on Social Presence in an Online Cooperative Game*. Springer International Publishing, Cham, 102–114. DOI: http://dx.doi.org/10.1007/978-3-319-46100-7_9

34. Joshua Newn, Eduardo Velloso, Fraser Allison, Yomna Abdelrahman, and Frank Vetere. 2017. 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 (CHI PLAY '17)*. ACM, New York, NY, USA, 541–552. DOI: http://dx.doi.org/10.1145/3116595.3116624

35. Joshua Newn, Eduardo Velloso, Marcus Carter, and Frank Vetere. 2016. Exploring the Effects of Gaze Awareness on Multiplayer Gameplay. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts (CHI PLAY Companion '16)*. ACM, New York, NY, USA, 239–244. DOI:http://dx.doi.org/10.1145/2968120.2987740

36. C. J. Van Rijsbergen. 1979. *Information Retrieval* (2nd ed.). Butterworth-Heinemann, Newton, MA, USA.

37. Katie Salen and Eric Zimmerman. 2003. *Rules of Play: Game Design Fundamentals*. The MIT Press.

38. Gita Sukthankar, Christopher Geib, Hung Hai Bui, David Pynadath, and Robert P. Goldman. 2014. *Plan, Activity, and Intent Recognition: Theory and Practice* (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

39. Jayson Turner, Eduardo Velloso, Hans Gellersen, and Veronica Sundstedt. 2014. EyePlay: Applications for Gaze in Games. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-human Interaction in Play (CHI PLAY '14)*. ACM, New York, NY, USA, 465–468. DOI: http://dx.doi.org/10.1145/2658537.2659016

40. Vincent van Rheden, Bernhard Maurer, Dorothé Smit, Martin Murer, and Manfred Tscheligi. 2017. LaserViz: Shared Gaze in the Co-Located Physical World. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction (TEI '17)*. ACM, New York, NY, USA, 191–196. DOI: http://dx.doi.org/10.1145/3024969.3025010

41. Eduardo Velloso and Marcus Carter. 2016. The Emergence of EyePlay: A Survey of Eye Interaction in Games. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '16)*. ACM, New York, NY, USA, 171–185. DOI: http://dx.doi.org/10.1145/2967934.2968084

42. Douglas Brent West and others. 2001. *Introduction to graph theory*. Vol. 2. Prentice hall Upper Saddle River.

A.3 Paper: *Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents*

DOI: 10.1007/978-3-030-29384-0_17

Designing Interactions with Intention-Aware Gaze-Enabled Artificial Agents

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,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 the counterparts.

Keywords: Human-AI interaction · Intention recognition · Explainable interface · Human-AI teaming · Intention awareness · Eye tracking · Gaze awareness

© IFIP International Federation for Information Processing 2019
Published by Springer Nature Switzerland AG 2019
D. Lamas et al. (Eds.): INTERACT 2019, LNCS 11747, pp. 255–281, 2019.
https://doi.org/10.1007/978-3-030-29384-0_17

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 [17,43,60]. The ability to tell what someone is looking at—'gaze awareness'—is a useful way to gauge the attention of others [1,2,14,63]. Gaze observed over time is an effective predictor of human intention [26,27,50,56]. 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 [50]. Despite these benefits, overlaying gaze on the interface can add a highly distracting element to the task at hand [50], 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 [27,56], 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 [14,50,51,63]. As nonverbal cues are challenging 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 the intentions from gaze input into written predictions through an artificial agent. We designed a within-subjects user study with three conditions, in which we provide varying levels of information to an assisted player. In the first condition, we provide players with a live 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 is 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. These findings suggest that the manner in which the agent communicates requires further exploration.

All in all, our work presents two primary contributions to the design of artificial agents that collaborate with humans, from both sides of the interaction. From the agent end, we show that it is possible to develop agents that can not only predict intentions through gaze but communicate and reason about them as well. On the other end, we show that the human counterpart can be supported by a proactive agent that communicates intentions through verbal means (e.g. written language), 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 [57]). These gaze behaviours provide clues about the other person's cognitive processes, i.e. the ability to discern their intentions [51]. The benefits are well demonstrated in multi-user scenarios, improving communication and coordination in collaborative settings (e.g. [2,12,24,63]). Gaze visualisation has also been explored in competitive gameplay [46,59], 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 [50,51].

Despite its numerous benefits, researchers have commonly found that using live gaze visualisation can be 'distracting' and 'confusing' for an observer to interpret [14,51,63]. 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 [50], 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 [15,50], 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 [50]. 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 alternative approach that incorporates visual behaviour into model-based intention recognition using automated planning [56]. 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 additional 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.

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

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 with regards 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 [30,48]. 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 [6,10,35,62].

Researchers in AI argue that providing explanations supports transparency and may improve trust in the system [23,41,47,52,62]. Moreover, when using an agent as a decision aid, users would often seek an explanation of its output to improve their own decision making [61]. 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, hence decreasing its effectiveness [52]. 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]. Licklider 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, 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 multi-user settings. Recent work in AI has shown that 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 an intelligent collaborative 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 Licklider's 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 trust of the user. Lastly, we need

to consider the optimum level of support as different levels of artificial agent 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, e.g. 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 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.

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

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 [50]. 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 and Procedure

Upon arrival, we sat the participant in front of a computer and obtained the participant's 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, 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 [50]. 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 of the study.

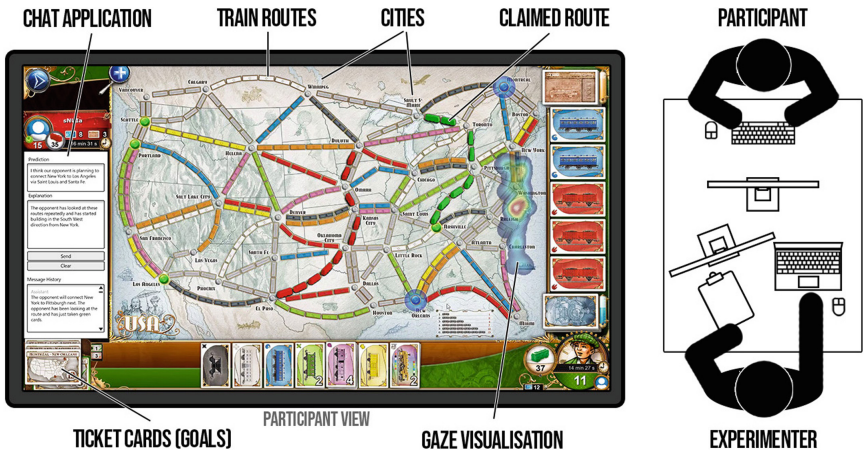


Fig. 1. Left: Participant view. **Right**: Experimental setup.

The recording of the game included a 'live' dynamic heatmap visualisation of the gaze of the 'opponent' player (as shown in Fig. 1-Left). We designed and employed a protocol to continually reinforce the participants' belief that they were engaged in a live online game 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 [32,54]. 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 [55].

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 step 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 'team-

mate' 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 Fig. 1-Left). 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 often 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", "Hmmm ok").

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, which allows 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 ($mean = 12.4$ 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 recommendations 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 and 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 Fig. 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 the opponent has performed only a few observable actions.

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

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 and 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 simply: intuition) 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 attached to the bottom of their screens; 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.

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.

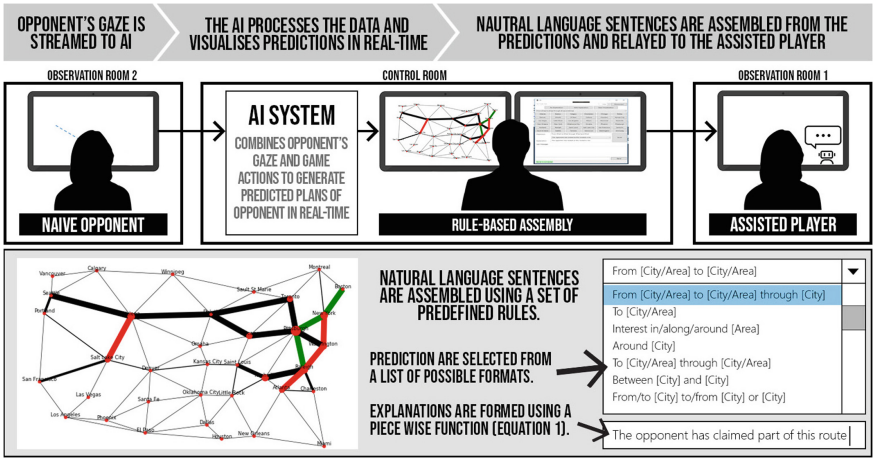


Fig. 2. Top: Experimental Setup and Communication Flow. **Bottom**: AI System Visualisation and Assembly Process. Opponent's intentions are displayed by increasing the line thickness of routes. The thicker the line, the more likely the route will be claimed. Coloured lines represent the claimed routes (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). (Color figure online)

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 [56]. 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 [49]. 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 from our previous study [50], 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 goals of the opponent (destination cities) to the wizard, the system remains far

better at discriminating and predicting the most likely plans and can provide this information earlier as well. This capability ensures relative consistency of predictions and provides a realistic impression of what such systems can do across participants. Third, as found in our previous study [50], 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 Fig. 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, i.e. alternate plans.

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 naive opponent (P_B) throughout the game using a live heatmap visualisation (as shown in Fig. 1-Left). This condition allowed players to make their inferences on their opponent's plans at the cost of their attention and serves as a baseline condition as we displayed the visualisation throughout the game.

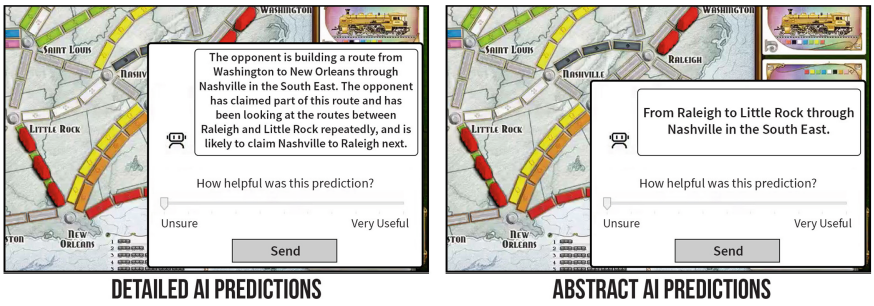


Fig. 3. AI prediction examples.

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 the strategies described in Sect. 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 a minute (60 seconds), 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, we only subjected the conditions to the assisted player (P_A), in which we presented as a 'mode of assistance'. For both AI 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 the predictions.

5.3 Measures and 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 intention awareness, we formed our questions based on our previous work on gaze-based intention recognition [50], 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 included questions to measure 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 additional 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 post-study semi-structured interview.

To measure cognitive workload unobtrusively, we used 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 question-

naire [22] to capture perceived workload based on six subscales—mental demand, physical demand, temporal demand, performance, effort and frustration.

5.4 Participants and Procedures

To manage the complexities of the study, all three researchers involved in the experiment 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, which will vary according to the condition.

In all conditions, the player received prompts with a slider (see Fig. 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 filling up the questionnaire. Once completed, the facilitator 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 player to the eye tracker before the tutorial. At the end of each condition, we administered the NASA-TLX questionnaire 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 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 previously outlined in Sect. 5.3 (Measures and Analysis). In the second part, we present and discuss the experience of the players with and without the agent from the 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 repeated-measures questionnaire.

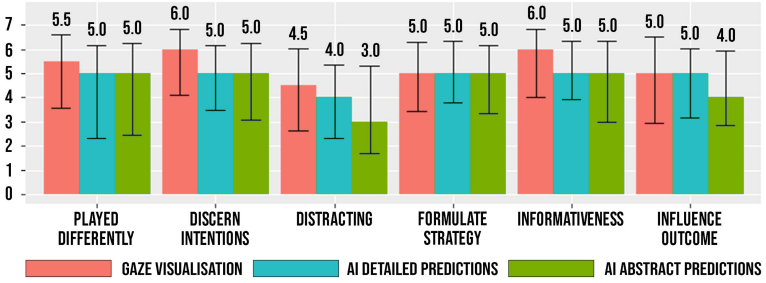


Fig. 4. Questionnaire results.

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.

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 the post-study semi-structured interview as players were asked to reflect 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 (>500 ms)	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 while achieving similar 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 questionnaire scores indicated that players perceived the 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 VIZ 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, $P5_A$ 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. From [City] to [City] in the 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 sent each time, it took them further away from their current task. Between the AI conditions, players preferred the abstract AI predictions over the detailed AI predictions in general as the messages were more concise and therefore needed less time investment in deciphering them and subsequently utilising the information:

- **$P8_A$**: *"I liked the simplicity of the information it [the artificial agent] gave me, it was very easy to filter."*
- **$P10_A$**: *"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."*
- **$P23_A$**: *"Shorter and brief hints were easy to understand and helpful."*

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 or when the plans of the opponent have changed, as $P15_A$

states *"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"*. Moreover, players also 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) at the same time, 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 their own, while $P12_A$ found it interesting to reaffirm their 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 Fig. 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 at a glance.

A few players drew comparisons with the AI predictions, for example, "*I prefer it [gaze visualisation] to the AI because I didn't have to bother with reading the pop-ups.*" as mentioned by $P23_A$. Players 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 additional information at all times in the GAZE VIZ condition. In comparison with the AI conditions, new information was only available when the predictions appeared, leaving the players on occasion to wait longer for new information to be sent.

6 Discussion

In this paper, we evaluated the prospects of an ideal intention-aware artificial agent, which we designed in line with the existing 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 players 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 same interface which could potentially clutter the interface and cause confusion. Due to the limitations of our approach concerning representation and context, we have only partially achieved our goals for a collaborative intention-aware artificial agent. Following, we discuss the considerations when designing such agents derived 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 participants 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 using 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 consider 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 therefore, better relevance to the user 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 player already knew, leading to the communication of redundant information. The subjective assessment of this condition shows that it is 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 improve the human's ability to respond to the agent. Furthermore, if the agent understood the intentions of all observable users, it would be possible for the agent to negotiate the broader goals of each of the users derived from their intentions. Our work closely resembles *iTourist*, in which an agent could recognise gaze patterns of a 'tourist' and provide recommendations on transport or accommodation alternatives [53], 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 scaled easily 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 [56], 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. [58]'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 their cell, and could 'see' that the human is consistently looking at a bench in a cell that was not their own. The robot then 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, improving its task efficiency and the interaction with the human. 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 [42,47], 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 possesses the ability to maintain the mental model of users with regards to short and long-term intentions that we can interrogate at any point in time using our 'white-box' approach. 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 various 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, including the presentation of information, the need for context-awareness, and opportunities in harnessing nonverbal communication.

Overall, the paper highlights the use of nonverbal behavioural inputs in Human-Agent Interaction and further provides an approach that can be applied in scenarios where it is important to know the intentions of others (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 to generate alternative predictions about another user based on 'what-if' queries (such as querying about an action that another user is most likely to take). These extended capabilities will be particularly useful in collaborative scenarios, where an agent can assist, mediate or negotiate with knowledge of multiple users' intentions.

Acknowledgements. We acknowledge the Australian Commonwealth Government, Microsoft Research Centre for Social Natural User Interfaces and Interaction Design Lab for their support on this project. We extend our gratitude to our colleagues, particularly Vassilis Kostakos, Tilman Dingler and Ryan Kelly for their input on the paper.

Dr. Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Award (Project Number: DE180100315) funded by the Australian Commonwealth Government.

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., Balkrishan, D.K., O'Neill, J., Winckler, M. (eds.) INTERACT 2017. LNCS, vol. 10514, pp. 447–466. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-67684-5_28
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, ETRA 2018, pp. 32:1–32:9. ACM, New York (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. LNCS, vol. 3205, pp. 54–69. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-30119-6_4
4. Argyle, M., Cook, M.: Gaze and Mutual Gaze. Cambridge University Press, Cambridge (1976)
5. Bednarik, R., Eivazi, S., Vrzakova, H.: Eye gaze in intelligent user interfaces. In: Nakano, Y., Conati, C., Bader, T. (eds.) A Computational Approach for Prediction of Problem-Solving Behavior Using Support Vector Machines and Eye-Tracking Data, pp. 111–134. Springer, London (2013). https://doi.org/10.1007/978-1-4471-4784-8_7
6. Biran, O., Cotton, C.: Explanation and justification in machine learning: a survey. In: IJCAI 2017 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, CHI EA 2018, pp. D504:1–D504:4. ACM, New York (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. J. Exp. Soc. Psychol. **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. LNCS (LNAI), vol. 8077, pp. 37–48. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40942-4_4
10. Chen, J., Barnes, M.: Human-agent teaming for multirobot control: a review of human factors issues. IEEE Trans. Hum. Mach. Syst. **44**(1), 13–29 (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. Theor. Issues Ergon. Sci. **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, CHI 2017, pp. 6245–6290. ACM, New York (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, ETRA 2019, pp. 79:1–79:5. ACM, New York (2019). https://doi.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, CHI 2016, pp. 2492–2496. ACM, New York (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, CHI 2018, pp. 349:1–349:12. ACM, New York (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, CHI 2018, pp. 562:1–562:12. ACM, New York (2018). https://doi.org/10.1145/3173574.3174136
17. Duchowski, A.T.: Gaze-based interaction: a 30 year retrospective. Comput. Graph. **73**, 59–69 (2018). https://doi.org/10.1016/j.cag.2018.04.002
18. Duchowski, A.T., et al.: 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, CHI 2018, pp. 282:1–282:13. ACM, New York (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, IUI 2018, pp. 211–223. ACM, New York (2018). https://doi.org/10.1145/3172944.3172961
20. Endsley, M.R.: Toward a theory of situation awareness in dynamic systems. Hum. Factors J. Hum. Factors Ergon. Soc. **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., Vilhjálmsson, H.H. (eds.) IVA 2009. LNCS (LNAI), vol. 5773, pp. 132–145. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-04380-2_17
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/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, HRI 2017, pp. 303–312. ACM, New York (2017). https://doi.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, CHI 2016, pp. 5180–5190. ACM, New York (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. Front. Psychol. **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, HRI 2016, pp. 83–90. IEEE Press, Piscataway (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, CHI 1990, pp. 11–18. ACM, New York (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.: The need for user models in generating expert system explanation. Int. J. Expert Syst. **1**(4), 345–375 (1988). http://dl.acm.org/citation.cfm?id=58447.58452
31. Keil, F.C.: Explanation and understanding. Annu. Rev. Psychol. **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 (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 Intell. Syst. **19**(6), 91–95 (2004). https://doi.org/10.1109/MIS.2004.74
34. Land, M.F.: Vision, eye movements, and natural behavior. Vis. Neurosci. **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, ACE 2016, pp. 17:1–17:9. ACM, New York (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 Trans. **87**(6), 1290–1298 (2004)
38. Licklider, J.C.R.: Man-computer symbiosis. IRE Trans. Hum. Factors Electron. HFE-1 **1**, 4–11 (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, UbiComp 2011, pp. 415–424. ACM, New York (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, AAMAS 2019, pp. 2432–2434. International Foundation for Autonomous Agents and Multiagent Systems, Richland (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, AAMAS 2019, pp. 1033–1041. International Foundation for Autonomous Agents and Multiagent Systems, Richland (2019). http://dl.acm.org/citation.cfm?id=3306127.3331801

43. Majaranta, P., Bulling, A.: Eye Tracking and eye-based human-computer interaction. In: Fairclough, S.H., Gilleade, K. (eds.) Advances in Physiological Computing. HIS, pp. 39–65. Springer, London (2014). https://doi.org/10.1007/978-1-4471-6392-3_3
44. Malle, B.F., Knobe, J.: Which behaviors do people explain? A basic actor-observer asymmetry. J. Pers. Soc. Psychol. **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. In: Wallner, G., Kriglstein, S., Hlavacs, H., Malaka, R., Lugmayr, A., Yang, H.-S. (eds.) ICEC 2016. LNCS, vol. 9926, pp. 102–114. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46100-7_9
46. Maurer, B., Lankes, M., Tscheligi, M.: Where the eyes meet: lessons learned from shared gaze-based interactions in cooperative and competitive online games. Entertainment Comput. **27**, 47–59 (2018). https://doi.org/10.1016/j.entcom.2018.02.009, http://www.sciencedirect.com/science/article/pii/S1875952117300629
47. Miller, T.: Explanation in artificial intelligence: insights from the social sciences. Artif. Intell. **267**, 1–38 (2019). https://doi.org/10.1016/j.artint.2018.07.007
48. Moore, J.D., Paris, C.L.: Requirements for an expert system explanation facility. Comput. Intell. **7**(4), 367–370 (1991)
49. 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, HRI 2015 Extended Abstracts, pp. 161–162. ACM, New York (2015). https://doi.org/10.1145/2701973.2702042
50. 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, CHI 2018, pp. 261:1–261:12. ACM, New York (2018). https://doi.org/10.1145/3173574.3173835
51. 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, CHI PLAY 2017, pp. 541–552. ACM, New York (2017). https://doi.org/10.1145/3116595.3116624
52. Nunes, I., Jannach, D.: A systematic review and taxonomy of explanations in decision support and recommender systems. User Model. User-Adap. Inter. **27**(3–5), 393–444 (2017). https://doi.org/10.1007/s11257-017-9195-0
53. 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, CHI 2005, pp. 221–230. ACM, New York (2005). https://doi.org/10.1145/1054972.1055004
54. Riek, L.D.: Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. J. Hum.-Robot Interact. **1**(1), 119–136 (2012). https://doi.org/10.5898/JHRI.1.1.Riek
55. Rieser, V.: Bootstrapping reinforcement learning-based dialogue strategies from wizard-of-Oz data. DFKI (2008)
56. 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 2018. International Foundation for Autonomous Agents and Multiagent Systems, Richland (2018). http://dl.acm.org/citation.cfm?id=3237383.3237457

57. 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, ICMI 2004, pp. 9–15. ACM, New York (2004). https://doi.org/10.1145/1027933.1027936
58. Unhelkar, V.V., et al.: Human-aware robotic assistant for collaborative assembly: integrating human motion prediction with planning in time. IEEE Robot. Autom. Lett. **3**(3), 2394–2401 (2018). https://doi.org/10.1109/LRA.2018.2812906
59. 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, CHI PLAY 2016, pp. 171–185. ACM, New York (2016). https://doi.org/10.1145/2967934.2968084
60. 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, CHI 1999, pp. 294–301. ACM, New York (1999). https://doi.org/10.1145/302979.303065
61. 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, CHI 2019, pp. 601:1–601:15. ACM, New York (2019)
62. 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)
63. Zhang, Y., Pfeuffer, K., Chong, M.K., Alexander, J., Bulling, A., Gellersen, H.: Look together: using gaze for assisting co-located collaborative search. Pers. Ubiquit. Comput. **21**(1), 173–186 (2017). https://doi.org/10.1007/s00779-016-0969-x

B

Study Materials

B.1 Study 2 Study Materials: Questionnaires

A

STUDY QUESTIONNAIRE

PROJECT TITLE:	Exploring Gaze in Immersive Gaming Environments
RESPONSIBLE RESEARCHER:	Eduardo Velloso
OTHER RESEARCHERS:	Joshua Newn, Fraser Allison & Frank Vetere

PARTICIPANT DETAILS

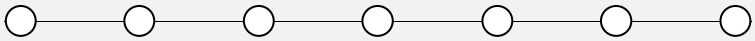
GAME NUMBER	AGE	GENDER
HOW MANY DAYS DID YOU PLAY GAMES IN THE LAST WEEK?		
WHAT KIND OF GAMES DO YOU PLAY?		
WHICH OF THESE BESTS DESCRIBES YOU: *NON-GAMER*, *OCCASIONAL GAMER*, *REGULAR GAMER*?		
HAVE YOU PLAYED TICKET TO RIDE PREVIOUSLY?		
HAVE YOU HAD ANY EXPERIENCE WITH EYE TRACKING?		

A

ROUND 1

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

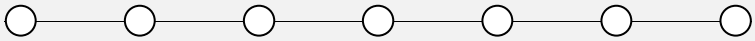
1) How well were you able to **predict** your **opponent's moves**?
Moves refer to the actions of your opponent e.g. claiming a route, picking a card, etc.



I was not able to predict at all.

I was able to predict extremely well.

2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.



I was not able to predict at all.

I was able to predict extremely well.

3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies**?

_____

_____

_____

_____

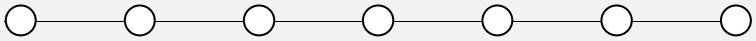
_____

A

ROUND 2

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) How well were you able to **predict** your **opponent's moves**?
Moves refer to the actions of your opponent e.g. claiming a route, picking a card, etc.



I was not able to predict at all.

I was able to predict extremely well.

2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.

I was not able to predict at all.

I was able to predict extremely well.

3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies?**

4) Did you find it **helpful** to know where your **opponent was looking** to find out **their strategy**?

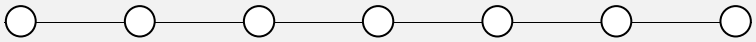
It was extremely unhelpful, and I found it to be distracting.

I did not find it helpful, but it did not bother me either.

It was extremely helpful, and I found it to be informative.

A

5) Did knowing where your opponent was looking make it **easier to predict their moves**?

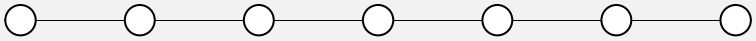


It did not make it easier to predict their moves at all.

It helped, but it did not make it significantly easier.

It made it extremely easy to predict their moves.

6) Did you find it **helpful** to know where your **opponent was looking** to formulate **your strategy**?

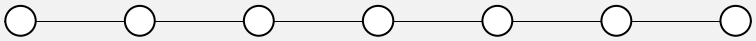


It was extremely unhelpful, and I found it to be distracting.

I did not find it helpful, but it did not bother me either.

It was extremely helpful, and I found it to be informative.

7) Did knowing where your opponent was looking **influence** the **outcome of the game**?

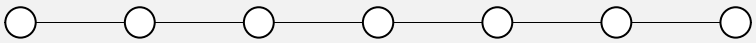


It did not have an influence on the game.

It had some influence but was not the deciding factor.

It had a clear influenced on the outcome of the game.

8) Did you **change the way you played** the game when you could see where you opponent was looking?



I did not change the way I played the game.

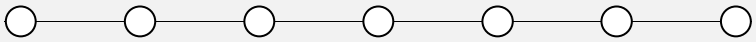
I completely changed the way I played the game.

A

ROUND 3

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) How well were you able to **predict** your **opponent's moves**?
Moves refer to the actions of your opponent e.g. claiming a route, picking a card, etc.



I was not able to predict at all.

I was able to predict extremely well.

2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.

I was not able to predict at all.

I was able to predict extremely well.

3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies**?

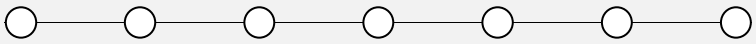
_____

_____

_____

_____

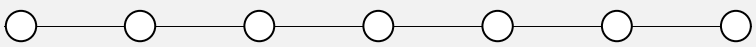
4) Did it **bother** you that your opponent knows where you were looking?



It did not bother me at all.

It bothered me a lot.

5) Did you **change the way you played** the game when your opponent could see where you were looking?



I did not change the way I played the game.

I completely changed the way I played the game.

A

6) Did you attempt to **deceive** your opponent? If so, how?

7) Did you attempt to **hide** your strategy? If so, how?

8) Additional comments/suggestions.

N

STUDY QUESTIONNAIRE

PROJECT TITLE:	Exploring Gaze in Immersive Gaming Environments
RESPONSIBLE RESEARCHER:	Eduardo Velloso
OTHER RESEARCHERS:	Joshua Newn, Fraser Allison & Frank Vetere

PARTICIPANT DETAILS

GAME NUMBER	AGE	GENDER
HOW MANY DAYS DID YOU PLAY GAMES IN THE LAST WEEK?		
WHAT KIND OF GAMES DO YOU PLAY?		
WHICH OF THESE BESTS DESCRIBES YOU: *NON-GAMER*, *OCCASIONAL GAMER*, *REGULAR GAMER*?		
HAVE YOU PLAYED TICKET TO RIDE PREVIOUSLY?		
HAVE YOU HAD ANY EXPERIENCE WITH EYE TRACKING?		

N

ROUND 1

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) How well were you able to **predict** your **opponent's moves**?
Moves refer to the actions of your opponent e.g. claiming a route, picking a card, etc.

I was not able to predict at all.

I was able to predict extremely well.

2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.

I was not able to predict at all.

I was able to predict extremely well.

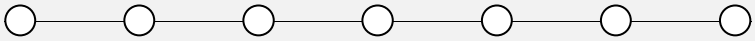
3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies**?

N

ROUND 2

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) How well were you able to **predict** your **opponent's moves**?
Moves refer to the actions of your opponent e.g. claiming a route, picking a card.



I was not able to predict at all.

I was able to predict extremely well.

2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.

I was not able to predict at all.

I was able to predict extremely well.

3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies**?

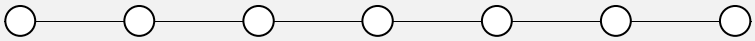
4) In what way was this round of the game **different** to the previous?

N

ROUND 3

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) How well were you able to **predict** your **opponent's moves**?
Opponent's moves refer to the actions of your opponent e.g. claiming a route, picking a card.



I was not able to predict at all.

I was able to predict extremely well.

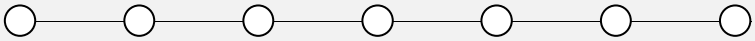
2) How well were you able to **predict** your **opponent's overall strategy**?
Overall strategies refer to the cities your opponent was attempting to connect, and the potential routes and actions they may have taken to get there.

I was not able to predict at all.

I was able to predict extremely well.

3) What **aspects** of the game did you find **helpful** in **predicting** your **opponent's strategies**?

4) Did you find it **helpful** to know where your **opponent was looking** to find out **their strategy**?



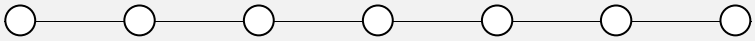
It was extremely unhelpful, and I found it to be distracting.

I did not find it helpful, but it did not bother me either.

It was extremely helpful, and I found it to be informative.

N

5) Did knowing where your opponent was looking make it **easier to predict their moves**?



It did not make it easier to predict their moves at all.

It helped, but it did not make it significantly easier.

It made it extremely easy to predict their moves.

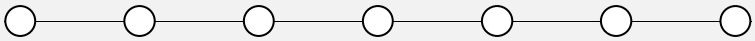
6) Did you find it **helpful** to know where your **opponent was looking** to formulate **your strategy**?

It was extremely unhelpful, and I found it to be distracting.

I did not find it helpful, but it did not bother me either.

It was extremely helpful, and I found it to be informative.

7) Did knowing where your opponent was looking **influence** the **outcome of the game**?

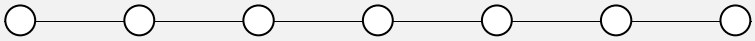


It did not have an influence on the game.

It had some influence but was not the deciding factor.

It had a clear influenced on the outcome of the game.

8) Did you **change the way you played** the game when you could see where your opponent was looking?



I did not change the way I played the game.

I completely changed the way I played the game.

9) Do you think whether your opponent was **aware** that they were being watched?

I did not think my opponent was aware.

I suspected but was not sure.

I was sure that my opponent was aware.

N

10) Did you feel you were **deceived** at any point in the game? If so, how?

11) Additional comments/suggestions.

B.2 Study 5 Study Materials: Questionnaires

A

STUDY QUESTIONNAIRE

PROJECT TITLE:	Exploring Gaze in Immersive Gaming Environments
RESPONSIBLE RESEARCHER:	Eduardo Velloso
OTHER RESEARCHERS:	Joshua Newn, Fraser Allison, Ronal Singh, Prashan Madumal, Frank Vetere

PARTICIPANT DETAILS

GAME NUMBER	AGE	GENDER
HOW MANY DAYS DID YOU PLAY GAMES IN THE LAST WEEK?		
WHAT KIND OF GAMES DO YOU PLAY?		
WHICH OF THESE BESTS DESCRIBES YOU: *NON-GAMER*, *OCCASIONAL GAMER*, *REGULAR GAMER*?		
HAVE YOU PLAYED TICKET TO RIDE PREVIOUSLY?		
HAVE YOU HAD ANY EXPERIENCE WITH EYE TRACKING?		

A

GAZE VISUALIZATION

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

A

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) Did you find it **helpful** to know where your **opponent was looking** to find out **their plans**?

It was extremely unhelpful. | It was neither helpful nor unhelpful. | It was extremely helpful.

2) Did you find it **helpful** to know where your **opponent was looking** to formulate **your own plans**?

It was extremely unhelpful. | It was neither helpful nor unhelpful. | It was extremely helpful.

3) Did knowing where your opponent was looking **influence** the **outcome of the game**?

It did not influence the outcome of the game. | It had some influence but was not the deciding factor. | It had a clear influenced on the outcome of the game.

4) Did you **change the way you played** the game when you could see where your opponent was looking?

I did not change the way I played the game. | I completely changed the way I played the game.

5) Did you find it **informative** to know where your opponent was looking?

It was not informative at all. | It was extremely informative.

A

6) Did you find it **distracting** to see where your opponent was looking?

It was not distracting at all.

It was extremely distracting.

7) Did the **visualization** of your opponent's gaze make it **difficult** to play the game?

It did not make it difficult to play the game.

It made it extremely difficult to play the game.

8) What aspects did you like or disliked when shown the gaze of your opponent?

9) In general, what did you think of the round you just played? (Interview Question)

A

AI PREDICTIONS

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

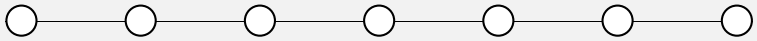
How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

A

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) Did you find it **helpful** that the AI predicted the intentions of your opponent?



It was extremely unhelpful. | It was neither helpful nor unhelpful. | It was extremely helpful.

2) Did you find it **helpful** to formulate your plans from knowledge of your opponent's intentions predicted by the AI?

It was extremely unhelpful. | I did not find it helpful. | It was extremely helpful.

3) Did being told your opponent's intentions **influence** the **outcome of the game**?

It did not influence the outcome of the game. | It had some influence but was not the deciding factor. | It had a strong influenced on the outcome of the game.

4) Did you **change the way you played** when the AI predicted your opponent's intentions?

I did not change the way I played the game. | I completely changed the way I played the game.

5) Overall, how **reliable** was the AI in predicting the opponent's intentions?

The AI was not reliable at all. | The AI was extremely reliable.

A

6) How well were you able to **keep track** of your opponent's intentions?

I was not able to keep track at all.

I was able to keep track at all times.

7) Did you find the AI predictions on your opponent's intentions **informative**?

It was not informative at all.

It was extremely informative.

8) Did you find the AI predictions on your opponent's intentions **distracting**?

It was not distracting at all.

It was extremely distracting.

9) How well did you **understand** the predictions provided from the AI?

I did not understand what the AI was telling me.

I completely understand what the AI was telling me.

10) What **aspects** did you like or disliked when given predictions by the AI?

_____

_____

_____

11) In general, what did you think of the round you just played? (Interview)

12) Did you find the AI helpful? (prompt)

A

AI EXPLAINED PREDICTIONS

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

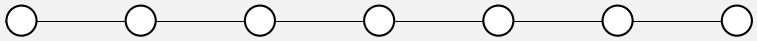
How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

A

The questions below are based on the round you just played. Please tick the circle that best describes your experience.

1) Did you find it **helpful** that the AI predicted the intentions of your opponent?



It was extremely unhelpful. | It was neither helpful nor unhelpful. | It was extremely helpful.

2) Did you find it **helpful** to formulate your plans from knowledge of your opponent's intentions predicted by the AI?

It was extremely unhelpful. | I did not find it helpful. | It was extremely helpful.

3) Did being told your opponent's intentions **influence** the **outcome of the game**?

It did not influence the outcome of the game. | It had some influence but was not the deciding factor. | It had a strong influenced on the outcome of the game.

4) Did you **change the way you played** when the AI predicted your opponent's intentions?

I did not change the way I played the game. | I completely changed the way I played the game.

5) Overall, how **reliable** was the AI in predicting the opponent's intentions?

The AI was not reliable at all. | The AI was extremely reliable.

A

6) How well were you able to **keep track** of your opponent's intentions?

I was not able to keep track at all. — I was able to keep track at all times.

7) Did you find the AI predictions on your opponent's intentions **informative**?

It was not informative at all. — It was extremely informative.

8) Did you find the AI predictions on your opponent's intentions **distracting**?

It was not distracting at all. — It was extremely distracting.

9) How well did you **understand** the predictions provided from the AI?

I did not understand what the AI was telling me. — I completely understand what the AI was telling me.

10) How **clear** did the AI provide the explained predictions?

They were not clear at all, and I required further clarification. — They were extremely clear, and I did not require further clarification.

11) What **aspects** did you like or disliked when given predictions by the AI?

_____

_____

12) In general, what did you think of the round you just played? (Interview)

A

OVERALL PREFERENCES

1) Which condition did you most prefer and least prefer?

Gaze Visualization	
AI Predictions	
AI Explained Predictions	

2) Which condition did you find the most useful and least useful?

Gaze Visualization	
AI Predictions	
AI Explained Predictions	

3) Which conditions did you find the most demanding and least demanding?

Gaze Visualization	
AI Predictions	
AI Explained Predictions	

4) Additional comments/suggestions.

B

STUDY QUESTIONNAIRE

PROJECT TITLE:	Exploring Gaze in Immersive Gaming Environments
RESPONSIBLE RESEARCHER:	Eduardo Velloso
OTHER RESEARCHERS:	Joshua Newn, Fraser Allison, Ronal Singh, Prashan Madumal, Frank Vetere

PARTICIPANT DETAILS

GAME NUMBER	AGE	GENDER
HOW MANY DAYS DID YOU PLAY GAMES IN THE LAST WEEK?		
WHAT KIND OF GAMES DO YOU PLAY?		
WHICH OF THESE BESTS DESCRIBES YOU: *NON-GAMER*, *OCCASIONAL GAMER*, *REGULAR GAMER*?		
HAVE YOU PLAYED TICKET TO RIDE PREVIOUSLY?		
HAVE YOU HAD ANY EXPERIENCE WITH EYE TRACKING?		

B

ROUND 1

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

B

Please indicate how you felt while playing the game for each of the items.

		Not at all	Slightly	Moderately	Fairly	Extremely
1	I felt content					
2	I felt skilful					
3	I thought it was fun					
4	I was fully occupied with the game					
5	I felt happy					
6	It gave me a bad mood					
7	I found it tiresome					
8	I felt competent					
9	I thought it was hard					
10	I forgot everything around me					
11	I felt good at it					
12	I felt bored					
13	I felt successful					
14	I enjoyed it					
15	I was fast at reaching the game's targets					
16	I felt annoyed					
16	I felt pressured					
17	I felt irritable					
18	I lost track of time					
19	I felt challenged					
20	I found it impressive					
21	I was deeply concentrated on the game					
22	I felt frustrated					
23	It felt like a rich experience					
24	I lost connection with the outside world					
25	I felt time pressure					
26	I had to put a lot of effort into it					

B

ROUND 2

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

B

Please indicate how you felt while playing the game for each of the items.

		Not at all	Slightly	Moderately	Fairly	Extremely
1	I felt content					
2	I felt skilful					
3	I thought it was fun					
4	I was fully occupied with the game					
5	I felt happy					
6	It gave me a bad mood					
7	I found it tiresome					
8	I felt competent					
9	I thought it was hard					
10	I forgot everything around me					
11	I felt good at it					
12	I felt bored					
13	I felt successful					
14	I enjoyed it					
15	I was fast at reaching the game's targets					
16	I felt annoyed					
16	I felt pressured					
17	I felt irritable					
18	I lost track of time					
19	I felt challenged					
20	I found it impressive					
21	I was deeply concentrated on the game					
22	I felt frustrated					
23	It felt like a rich experience					
24	I lost connection with the outside world					
25	I felt time pressure					
26	I had to put a lot of effort into it					

B

ROUND 3

Place a mark at the desired point on each scale that best indicates your experience for the round you just played.

Mental Demand

How mentally demanding was the task?

Very Low — Very High

Physical Demand

How physically demanding was the task?

Very Low — Very High

Temporal Demand

How hurried or rushed was the pace of the task?

Very Low — Very High

Performance

How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort

How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration

How insecure, discouraged, irritated, stressed and annoyed were you?

Low — High

B

Please indicate how you felt while playing the game for each of the items.

		Not at all	Slightly	Moderately	Fairly	Extremely
1	I felt content					
2	I felt skilful					
3	I thought it was fun					
4	I was fully occupied with the game					
5	I felt happy					
6	It gave me a bad mood					
7	I found it tiresome					
8	I felt competent					
9	I thought it was hard					
10	I forgot everything around me					
11	I felt good at it					
12	I felt bored					
13	I felt successful					
14	I enjoyed it					
15	I was fast at reaching the game's targets					
16	I felt annoyed					
16	I felt pressured					
17	I felt irritable					
18	I lost track of time					
19	I felt challenged					
20	I found it impressive					
21	I was deeply concentrated on the game					
22	I felt frustrated					
23	It felt like a rich experience					
24	I lost connection with the outside world					
25	I felt time pressure					
26	I had to put a lot of effort into it					