Human-Dog Visual Interaction

By

Madeline H. Pelgrim

H.B.Sc., University of Toronto, 2019

M.S., Brown University, 2023

Thesis

Submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in the Department of Cognitive and Psychological Sciences at Brown University

PROVIDENCE, RHODE ISLAND

May 2026

This dissertation by Madeline H. Pelgrim is accepted in its present form by the Department of Cognitive and Psychological Sciences as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Daphna Buchsbaum, Advisor

Recommended to the Graduate Council

Date _____ _____

Bertram Malle, Reader

Date _____ _____

David Sobel, Reader

Date _____ _____

Stefanie Tellex, Reader

Approved by the Graduate Council

Date _____ _____

David Lindstrom, Dean of the Graduate School

Madeline H. Pelgrim

Email: mpelgrim@gmail.com
Website: https://mpelgrim.github.io

Last Updated: April 20, 2026

Academic Appointments

Stonehill College — Easton, MA
Professor of Practice — *(exp) July 2026*

Education

Brown University — Providence, RI
Doctor of Philosophy in Cognitive Science — *September 2020 - May 2026 (exp.)*

Brown University — Providence, RI
Master of Science in Cognitive Science — *September 2020 - May 2023*

University of Toronto — Toronto, ON, Canada
Honors Bachelor of Science in Psychology & Biology with High Distinction — *September 2015 - June 2019*

Journal Publications

*Undergraduate mentees indicated with underline

Pelgrim, M. H., Sundara Raman, S., Serre, T., & Buchsbaum, D. (2025) Evaluating Dogs' Real World Visual Environment and Attention. *Cognitive Science*, 49 (6), e70080

Pelgrim, M. H., Tidd, Z., Byrne, M, Johnston, A. M., & Buchsbaum, D. (2024) Synchronous citizen science with dogs. *Animal Cognition*, 27(1): 46.

ManyDogs Project, Espinosa, J., Stevens, J.R., Alberghina, D., Alway, H.E.E., Barela, J.D., Bogese, M., Bray, E.E., Buchsbaum, D. Byosiere, S.-E., Byrne, M., Cavalli, C. M., Chaudoir, L. M., Collins-Pisano, C., DeBoer, H.J., Douglas L.E.L.C., Dror, S., Dzik, M.V., Ferguson, B., Fisher, L., Fitzpatrick, H.C., Freeman, M.S., Frinton, S.N., Glover, M.K., Gnanadesikan, G.E., Goacher, J.E.P., Golańska, M., Guran, C.-N.A., Hare, E., Hare, B. Hickey, M., Horschler, D.J., Huber, L., Jim, H.-L., Johnston, A.M., Kaminski, J. Kelly, D.M., Kuhlmeier, V.A., Lassiter, L., Lazarowski, L., Leighton-Birch, J., MacLean, E.L., Maliszewska, K., Marra, V., Montgomery, L.I., Murray, M.S., Nelson, E.K., Ostojić, L., Palermo, S.G., Parks Russell, A.E., **Pelgrim, M.H.**, Pellowe, S.D., Reinholz, A., Rial, L.A., Richards, E.M., Ross, M.A., Rothkoff, L.G., Salomons, H., Sanger, J.K., Santos, L., Shirle, A.R., Shearer, S.J., Silver, Z.A., Silverman, J.M., Sommese, A., Srdoc, T., St. John-Mosse, H., Vega, A.C., Vékony, K., Völter, C.J., Walsh C.J., Worth, Y.A., Zipperling, L.M.I., Żołędziewska, B., & Zylberfuden, S.G. (2023) ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension. *Animal Behavior & Cognition*, 10(3), 232-28.

ManyDogs Project, Alberghina, D., Bray, E.E., Buchsbaum, D., Byosiere, S.E., Espinosa, J., Gnanadesikan, G.E., Guran, A., Hare, E., Horschler, D.J., Huber, L., Kuhlmeier, V.A., MacLean, E.L., **Pelgrim, M.H.**, Perez, B., Ravid-Schurr, D., Rothkoff, L., Sexton, C.L., Silver, Z.A., & Stevens, J.R. (2023) ManyDogs Project: A Big Team Science Approach to Investigating Canine Behavior and Cognition. *Comparative Cognition & Behavior Reviews* 18, 59-77

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022) Head-Mounted Mobile Eye-tracking in the Domestic Dog: A New Method. *Behavior Research Methods* 55, 1924-1941

Pelgrim, M. H., Espinosa, J., Tecwyn, E.C., Marton, S. M., Johnston, A.M., & Buchsbaum, D. (2021) What's the point? Domestic Dogs' understanding of the accuracy of human social informants. *Animal Cognition* 24, 281-297.

Peer-Reviewed Conference Proceedings

Pelgrim M. H, He, I.X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024). Find it like a dog: Using Gesture to Improve Object Search *Proceedings of the 46th Annual Meeting of the Cognitive Science Society*

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2023) The Impact of Quality and Familiarity on Dogs' Food Preferences. *Proceedings of the 45th Annual Meeting of the Cognitive Science Society*

Pelgrim, M. H., & Buchsbaum, D. (2022) Categorizing Dogs' Real World Visual Statistics. *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

Conference Presentations

Pelgrim, M. H., He, I.X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024, Jul) Find it like a dog: Using Gesture to Improve Object Search. Talk at the *46th Annual Meeting of the Cognitive Science Society* in Rotterdam, Netherlands

Ross, M., **Pelgrim, M. H.**, Malle, B., & Buchsbaum, D. (2024, Jun) What makes Lassie smart? Human Perceptions of Canine Intelligence. *Talk at the Canine Science Conference* in Seattle, WA, USA

Pelgrim, M. H. & Buchsbaum, D. (2023, Apr) Categorizing Dogs' Real World Visual Statistics. Talk at the *30th Annual International Conference of the Comparative Cognition Society* in Melbourne Beach, FL, USA

Pelgrim, M. & Buchsbaum, D. (2022, Oct) Categorizing Dogs' Real World Visual Statistics. Talk at the *Canine Science Conference* in Hamilton, NY, USA

Pelgrim, M. & Buchsbaum, D. (2022, Jul) Categorizing Dogs' Real World Visual Statistics. Talk at the *44th Annual Conference of the Cognitive Science Society* online and in Toronto, ON, Canada

Pelgrim, M., Espinosa, J., Buchsbaum, D. (2022, Apr). Impact of Value and Familiarity on Dogs' Food Preferences. Talk at the *29th Annual International Conference of the Comparative Cognition Society* online

Pelgrim, M., Espinosa, J., Buchsbaum, D. (2021, Jul). Head Mounted Eye-Tracking In Dogs: A New Method. Talk at the *7th Biannual Canine Science Forum* online & in Lisbon, Portugal

ManyDogs, Espinosa, J., **Pelgrim, M.**, Byosiere S. (2021, Jul). The ManyDogs Project. Virtual round table at the *7th Biannual Canine Science Forum* online & in Lisbon, Portugal

Espinosa, J., Corbit, L., McGinn, K., **Pelgrim, M.**, Buchsbaum, D. (2021, Jul). Domestic dogs understanding of the temporal priority principle. Virtual Talk at the *7th Biannual Canine Science Forum* online & in Lisbon, Portugal

Pelgrim, M., Espinosa, J., Buchsbaum, D. (2021, Apr). Head Mounted Eye-Tracking In Dogs: An Update and Future Directions. Talk at the *East Coast Canine Cognition Workshop* online

Pelgrim, M., Espinosa, J., Buchsbaum, D. (2021, Apr). Wearable Eye-tracking in the Domestic Dog: A New Method. Talk at the *28th Annual International Conference of the Comparative Cognition Society* online

Gnanadesikan, Gitanjali E., Julia Espinosa & **ManyDogs**. (2021, Jan). ManyDogs 1: An International Collaborative Approach to Pointing Comprehension in Domestic Dogs. Presented at the *Animal Behaviour Twitter Conference.*

Espinosa, J., Corbit, L., McGinn, K., **Pelgrim, M.**, Buchsbaum, D. (2020, Sep). Exploring domestic dogs use of temporal directionality for inferring hidden causal relations. Talk presented virtually at the *University of Toronto's Ebbinghabus Empire Meeting* online.

Espinosa, J., Corbit, L., McGinn, K., **Pelgrim, M.**, Buchsbaum, D. (2020, Jul). Investigating the spatiotemporal priority principle in domestic dogs. Talk presented virtually at the *Animal Behaviour Society Virtual Meeting* online, originally scheduled to be in Knoxville, TN, USA

Pelgrim, M., Espinosa, J., Buchsbaum, D. (2020, Feb). Wearable Eye-tracking in the Domestic Dog: A New Method. Talk at the *East Coast Canine Cognition Workshop* in New Haven, CT, USA

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2019, Oct). Domestic Dogs' understanding of the accuracy of human social informants. Talk at the *Canine Science Conference* in Phoenix, AZ, USA

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2019, Jan). What's the point? Domestic dogs understanding of the accuracy of human social cues. Talk at the *Arts and Sciences Student Union Undergraduate Research Conference* in Toronto, ON, Canada.

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2018, Nov). Domestic dogs understanding of the accuracy of human social cues. Talk at the *East Coast Canine Cognition Workshop* in New Haven, CT, USA.

Espinosa, J., Tecwyn, E., **Pelgrim, M.**, Marton, S., Johnston, A., Buchsbaum, D. (2018, Jul). What's that point all about? Domestic dogs compare the accuracy of human social cues. Talk at the *Canine Science Forum* in Budapest, Hungary.

Conference Posters

Julia S. Ceccarelli, Carly F. Fisher, **Madeline H. Pelgrim**, & Daphna Buchsbaum. (2025, Jul). On a Roll: Physical Reasoning and the Understanding of Solidity in Dogs. Poster at the *Brown University Summer Research Symposium* in Providence, RI, USA

Pelgrim, M.H., He, I.X., Tellex, S., & Buchsbaum, D. (2025, Mar) Point comprehension in dogs and toddlers. Poster at the *33rd Annual International Conference of the Comparative Cognition Society* in Albuquerque, NM, USA

Pelgrim, M.H., He, I.X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024, Jun) What's in a point? Quantifying dog guardian pointing gestures. Poster at the *Canine Science Conference* in Seattle, WA, USA

Pelgrim, M.H., Espinosa, J., Buchsbaum, D. (2023, Jul) The Impact of Value and Familiarity on Dogs' Food Preferences. Poster at the *45th Annual Conference of the Cognitive Science Society* online & in Sydney, Australia

Gatesy-Davis, A., **Pelgrim, M.H.**, Buchsbaum, D. (2023, Apr). A Comparison of human trust in human and non-human agents. Poster at the *30th Annual International Conference of the Comparative Cognition Society* in Melbourne Beach, FL, USA

Ross, M., **Pelgrim, M. H.**, Buchsbaum, D. (2023, Apr). Are Dog Owner's Perceptions of their Pets' Intelligence Accurate? Poster at the *30th Annual International Conference of the Comparative Cognition Society* in Melbourne Beach, FL, USA

Gatesy-Davis, A., **Pelgrim, M.H.**, Buchsbaum, D. (2022, Oct). Measuring Humans' Trust in Dogs with the Multi-Dimensional Measure of Trust (MDMT). Poster at the *Canine Science Conference* in Hamilton, NY, USA

Gatesy-Davis, A., **Pelgrim, M.H.**, Buchsbaum, D. (2022, Apr). Measuring Humans' Trust in Dogs with the Multi-Dimensional Measure of Trust (MDMT). Poster at the *29th Annual International Conference of the Comparative Cognition Society* online

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2021, Apr). Domestic Dogs' understanding of the accuracy of human social informants. Poster at the *Society for Research in Child Development Biennial Conference* online

Espinosa, J., Corbit, L., McGinn, K., **Pelgrim, M.**, Buchsbaum, D. (2020, Jul). Domestic dogs' understanding of spatial temporal priority. Poster presented virtually at the *42nd Annual Meeting of the Cognitive Science Society* online

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2019, Jul). Domestic Dogs' understanding of the accuracy of human social informants. Poster presented at the *41st Annual Meeting of the Cognitive Science Society* in Montreal, QC, Canada

Pelgrim, M., Espinosa, J., Tecwyn, E., Marton, S., Johnston, A., Buchsbaum, D. (2019, Apr). Domestic Dogs' understanding of the accuracy of human social informants. Poster presented at the *26th Annual International Conference of the Comparative Cognition Society* in Melbourne Beach, FL, USA

Pelgrim, M., Espinosa, J., Tecwyn, E., Buchsbaum, D. (2017, Mar). Physical reasoning in the domestic dog: An exploration of solidity. *Undergraduate Research Opportunities Program Poster Session*, Toronto, ON, Canada

Teaching

Instructor

Cognitive Psychology, *Rhode Island School of Design.* Spring 2026.

Introduction to Psychology, *Bryant University.* Fall 2025.

Psychology Across Species: Animal Cognition & Behavior, *Brown University Pre-College Program*, Summer 2022 - 2025.

Learning & Cognition, *Providence College*, Fall 2023.

Supervisory Teaching Assistant

Mind, Brain and Behavior: An Interdisciplinary Approach, *Brown University*, Fall 2022.

Teaching Assistant

Research Methods, *Brown University*, Spring 2023.

Children's Thinking: Asynchronous, *Brown University*, Summer 2022, 2023.

Children's Thinking, *Brown University*, Spring 2022.

Mind, Brain and Behavior: An Interdisciplinary Approach, *Brown University*, Fall 2021.

Guest Lectures

Statistical Methods *Brown University*, Fall 2024
Children's Thinking *Brown University*, Spring 2022, 2024
Senior Seminar on Canine Cognition *Occidental College*, Spring 2024
Research Methods *Brown University*, Spring 2023
Topics in Development: The Developmental and Evolutionary Origins of Mind *Brown University*, Fall 2022

Selected Professional Development

Center for the Integration of Research, Teaching, and Learning (CIRTL): Associate Level, Certificate, An Introduction to Evidence-Based Undergraduate STEM Teaching, Summer 2024 8 week course completed with distinction; Change Leadership for Inclusive Teaching and Learning, Summer 2024 5 week course.

Brown University Sheridan Center for Teaching and Learning: Teaching Consultant Program, Fall 2024; Reflective Teaching Seminar Program, Fall 2022; Essentials for Graduate TA's, Fall 2021.

External Conferences: Teaching at Teaching Intensive Institutions Conference, 2024

Mentorship

Senior Honors Thesis Mentor Brown University

- Tovi Portnow (2026) - *"What does eye-tracking reveal about how toddlers use other's pointing cues when finding a hidden object?"*
- Julia S. Ceccarelli (2026) - *"Comparing Understanding of Solidity in Dogs and Dingoes"*
- Lauren Bonner (2024) - *"The Dog Days of Life: An Investigation into Factors that Impact Age-Related Cognitive and Behavioral Changes in Dogs"*
- Sarah Zylberfuden (2022) *"Do Dogs Understand Pointing as a Social Communicative Gesture"*

Research Mentor *Brown Dog Lab* Brown University, 2020 - Present. Mentored undergraduate students in research practice, coding, and data analysis, and provided support with professional development. Includes 4 honors thesis students, 1 independent study student (2022), and 13 students funded through Brown's competitive Undergraduate Teaching and Research Award program.

Peer Mentor *International Student Orientation* Brown University, 2021-2025. Served as peer mentor to approximately 20 incoming international graduate students during orientation each year.

Research Mentor *Student Research Institute by the Harvard Undergraduate OpenBio Laboratory.* Summer 2025. Mentored an advanced high school student in conducting an independent research project.

Leadership Alliance Mentor *Carney Institute for Brain Science* Brown University. 2024. Provided research training and professional development mentorship to 1 student per year from an underrepresented background and from a historically Black college and universities or minority-serving institutions.

Service & Leadership

Undergraduate Curriculum Committee *CLPS Department, Brown University* 2024 - Present. Elected as departmental graduate student representative to serve on the undergraduate curriculum committee to discuss coursework, honors requirements, curriculum needs, and majors offered by the department.

Committee Member *CLPS Department Diversity and Inclusion Action Plan, Brown University* 2022-Present. Work on Curriculum Subcommittee to assess and improve the curriculum of CLPS Department courses through planning reading groups, syllabi surveys, and developing new courses.

Primary Scientific Member *Brown University Institutional Animal Care and Use Committee* 2022 - Present. Conduct evaluations of animal care and use program, supervise and asses animal care and use activities, review and approve of institutional training programs

Assistant Director: Methods Development & Monitoring, *ManyDogs* 2022 - 2025. Leads committee to support ManyDogs projects implementing best practices in canine science research methods. Acts as spokesperson for the committee and provides regular updates to the Co-Directors about progress and problems.

Animal Research Compliance Advisory Group Member *Office of Research Integrity, Brown University*, 2021 - 2024. Collaborated and partner with animal research program stakeholders to promote positive animal welfare initiatives.

Graduate Representative *CLPS Department* Brown University, 2021-2022 Academic Year. Elected as one of two departmental representatives of graduate student body. Serve to bring student ideas and concerns to faculty and plan events like town-hall meetings, socials, and professional development seminars.

Meeting Organizer *Greater Toronto Area Animal Cognition Group* 2019 - 2020. Contacted and invited speakers from a wide range of disciplines and scheduled monthly meetings. Facilitated meetings, introducing speakers and moderating timing of discussion

Journal Reviewer *Animal Cognition, Developmental Science, Proceedings of the Royal Society B, Open Mind*

Invited Talks & Panels

Queens University, Social Cognition Lab, September 2025 *Cooperation Across Species*

University of California, Berkeley, Social Origins Lab, October 2024 *Find it Like a Dog*

Brown University Developmental Brown Bag, March 2024 *Evaluating Mechanisms of Selective Social Learning*

Brown University Departmental Undergraduate Group, November 2023, *Graduate School Panel*

Boston College Canine Cognition Center, March 2023. *What's that Dog Looking At?*

University of Toronto Student Recruitment Training, September 2019, *Introduction to Canine Cognition.*

University of Toronto International Counsellors Event, May 2019, *Research in Canine Cognition*

University of Toronto Psychology Student Association, November 2018, *Undergraduate Research*

Honors and Funding

NSERC Post-Graduate Scholarship - Doctoral, 2023-2026

Kuczaj Memorial Travel Grant, 2023. Comparative Cognition Society.

NIH Interdisciplinary Training in Computational, Cognitive, and Systems Neuroscience T32, 2022-2023.

NSF Graduate Research Fellowship Program Honorable Mention, 2022.

Mary L. S. Downes Fund Fellowship, 2021-2022. Brown University.

Graduate Conference Travel Award, 2021, 2022, 2024. Brown University.

Betty R.H. and James M. Pickett Fellowship Fund for Psychology, 2020-2021. Brown University.

Arts and Science Student Union Travel Grant, 2019. University of Toronto.

NSERC Undergraduate Student Research Award, 2018.

Selected Media Coverage

How Do Dogs See The World? They Do See Color, But They Focus More on Us (October, 2025) Discover Magazine

"What's Fido Thinking? Brown's canine cognition lab is not just a walk in the park." (October, 2024) Brown University Alumni Magazine.

"In Photos: U of T's Canine Cognition Lab, where dogs play and humans learn" (March, 2019) University of Toronto News.

"What does a Scientist Look Like?" (March, 2019) The Varsity, University of Toronto's Student Newspaper

"How do dogs learn? U of T undergrads look for answers"(March, 2017) University of Toronto News.

Acknowledgments

Before starting graduate school I was told that a PhD is a marathon and not a sprint. While certainly an appropriate metaphor, a marathon implies a solo pursuit. My journey through graduate school has been anything but, and has shown me support and love every step of the way. I am immensely grateful for those who have carried me across this finish line, it would not have been possible without you.

My advisor Daphna Buchsbaum, thank you for the support dating back to undergrad. You've been a supportive and understanding mentor, always encouraging me to strive for more. You have taught me so much about what it means to be a leader and a scientist.

My dissertation committee members, Dave Sobel, Bertram Malle, and Stefanie Tellex. You have each been incredibly generous with your time, knowledge, and your ideas. Thank you for your consistent feedback, support, and inspiration.

The small army of undergraduate research assistants - thank you. I have learned more from working with you than I think I have taught, but it has been an absolute privilege to know each of you. Particular thanks to Renee Kim and Amelia Wieland for their tireless efforts in coding on exceptionally short notice.

My covid-cohort mates Anthony Bruno and Kei Yoshida. Starting graduate school during a global pandemic was not ideal. Seeing you (and our friendly covid testers) was my only human contact the first semester and kept me sane. I will always cherish our weekly coffees, reading groups, and office panics.

My lab-mate and friend Carly Fisher. Self-titled equal contributor to all of my work - thank you for the coffees, the drinks, and the hours of editing. Your show and tell has always been a highlight and the ability to talk dogs and workshop problems helped me get through these last two years.

The dogs and humans in all of my studies - thank you. The dogs were all very good.

My friends near and far - sharing this season of life with each of you has been a joy. Alana, Haley, Kei, Sarah - the finest citizens this department has ever known. Rachel for the endless podcasts, Tom and Cheyenne for reminding me life continues outside of this department. Bookclub - for everything.

My dog Lilly and cat Fabio - thank you for listening to every talk I have ever given. You've been a terribly unreceptive audience but I have loved you regardless.

During my PhD I formally gained a set of parents - to my family through marriage and birth I could not have gotten here without your support. Thank you to Roger and Lisa for your unwavering love and positivity. Mom - thank you for your sacrifices and persistent belief in me even when I doubted myself.

Nick - there are no words that can explain how much better you've made my life. Thank you for every moment of joy and love you've brought to my life.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

What makes human beings unique? We can learn about ourselves by studying other species. Dogs provide a valuable opportunity to explore social learning and non-human cognition more broadly (Bensky et al., 2013). As a species, they perform critical roles in our society. In a working context, dogs help blind people navigate the world as guide dogs, detect explosives to keep air travel safe, and search disaster zones for missing people. As pets, dogs provide numerous health benefits to their owners (Hayden-Evans et al., 2018; Stanley et al., 2014). Despite our reliance on dogs for companionship and as service partners, we know little about how dogs visually perceive and engage with the world around them and with their human companions.

Dogs were domesticated thousands of years ago from ancient wolves (Germonpré et al., 2012; Thalmann et al., 2013). The number of domestication events and the specific driving factors that led to domestication remain controversial. However, there is broad consensus that dogs and humans have been co-existing for over 15,000 years. Unlike their wild ancestors, dogs are comfortable working with humans, even those they do not know. Although cultural differences exist in how people raise and view the dogs they live alongside, domestic dogs exist on every continent except Antarctica (Gray & Young, 2011; Wan et al., 2009).

Dogs' unique evolutionary relationship and close living arrangements with humans have resulted in dogs having an exceptional ability to respond to human social-communicative gestures. Much of the current work on canine cognition has focused on dogs' ability to follow human pointing gestures. These studies focus on dogs following a human point to find hidden food in a variety of contexts. In developmental paradigms, puppies are able to follow points from the earliest age that they can be tested (Agnetta et al., 2000; Hare et al., 2002). Dog pups also outperform wolf pups at following subtle human pointing gestures, even when the wolf pups have been hand raised and intensively socialized with humans (Virányi et al., 2008).

Dogs also do not follow blindly. They can selectively learn from human informants and integrate their own evidence with the information provided. Dogs are sensitive to the knowledge of an informant, choosing the location pointed to by a knower more often than the location pointed to by a guesser (Catala et al., 2017; Maginnity & Grace, 2014). They are also sensitive to the past accuracy of an informant. This is an ability they appear to share with human children. Dogs are able to use past informant accuracy to selectively trust, choosing the location indicated by a previously accurate individual significantly more than the location pointed to by a previously inaccurate person (Corriveau et al., 2009; Pelgrim et al., 2021).

Some argue that through domestication and artificial selection, the natural social partner for the dog has become the human rather than other dogs (Marshall-Pescini et al., 2013; Naderi et al., 2001; Range et al., 2019). In addition to the wealth of working and service roles dogs perform, dogs can recruit human partners to complete tasks requiring two individuals. A classic test of collaboration is the string pulling task, where a desirable item is placed on a board out of reach. The board has a string looped around the edges, and the string is within reach of the participants. To draw the item close, two individuals must pull on the string simultaneously. If only one individual pulls on the string, it will come loose from the board, and the item will remain inaccessible. When paired with another dog as their collaborative partner, dogs rarely succeed. They fail to coordinate and pull on the string simultaneously with other dogs. However, when paired with a human partner, dogs can successfully cooperate with the person to pull on

the string (Range et al., 2019). Not only can dogs complete string pulling tasks with a human partner, but dogs can also wait to pull the string until their partner arrives. To briefly summarize, dogs' evolutionary history and their typical living environment make them a promising system for studying social learning and interaction (Kubinyi et al., 2009).

1.1 Visual Interaction and Attention

Humans, like other primates, are a visual species. Many of our daily social interactions involve visual coordination and attention (Colavita, 1974; Ellsworth & Ludwig, 1972; Kaas, 1992; Kaas & Balaram, 2014). The importance of visual coordination is reflected in major theories of human uniqueness, such as the shared intentionality hypothesis (Tomasello et al., 2005). The shared intentionality hypothesis emphasizes the importance of engagement between individuals and characterizes this engagement in three ways, each with increasing complexity.

The simplest, dyadic engagement, involves two individuals sharing, interacting, or engaging in turn-taking. This emerges in human children during infancy and is exemplified by mutual gazing (where the infant and their partner look into each other's eyes). Dyadic engagement does not require the individual to represent or read goals or intentions. The individual may have these, but they do not require a representation of the partner's mental state. Dyadic engagement and the associated key visual behavioral signature (mutual gaze) is theorized to be an evolutionarily ancient ability that is shared with other species (Tomasello et al., 2005).

Next, children develop triadic engagement, which involves goal sharing and the monitoring of both the goal state. Children also have to integrate their own perception and the perception of their partner. Triadic engagement emerges in 9-12-month-old children. Critically, although some studies have shown that non-human primates are capable of triadic engagement, they engage in this style of interaction less frequently than human children. An example of a way to measure triadic engagement is re-engagement in an activity after a partner stops engaging (i.e., if the partner disengages from the game the child may hand the partner the toy or attempt

to communicate with the partner to entice them back to the game). The key visual behavioral signature (visual re-engagement) should only be fully displayed in human beings.

The final engagement style is collaborative engagement, which involves forming a coordinated action plan with an intentional social partner and then using joint attention to the task at hand to accomplish the shared goal. Under the shared intentionality hypothesis, this type of engagement is entirely unique to humans. Collaborative engagement emerges in children around 12 to 15 months of age. This style of engagement is classically identified by joint attention. Critically, joint attention has a specific visual behavioral signature that is taken as a proxy, or at least as a correlate, for the rich set of associated cognitive abilities. Joint attention involves both individuals carefully monitoring and representing the goals, actions, and intentions of others. When we engage in joint attention, we not only attend to a real-world item but we monitor our partner's attention. We have intentions that underlie our actions, and when engaging in joint attention we share those intentions with our partner, facilitating our cooperation and collaboration (Baldwin et al., 2001; Carpenter et al., 1995; Kaplan & Hafner, 2006; Malle, 2008).

Joint attention as a cognitive process is rich, highly social, and theoretically unique to our species. This cognitive process helps us to learn language, facilitating our comprehension and production of spoken language (Baldwin, 1995; Carpenter et al., 1998). At older ages, joint attention helps us to understand written language, with increased rates of joint attention corresponding to better written word learning (Guo & Feng, 2013).

Joint attention as a cognitive process is theorized to be unique to human beings, and may be necessary and sufficient to explain all of human unique cognition. It follows that the key visual behavioral signature of joint attention should also be unique to human beings. Much of the support for the shared intentionality hypothesis has been taken from behavioral observations of visual interaction. Prior work has focused on comparing the ability of chimpanzees and children to respond to human visual communication, and consistently finds evidence for joint attention in children but not in chimpanzees (Carpenter et al., 1995; Tomasello et al., 2005).

However, there has been much debate about the operationalization of joint attention (Bradley et al., 2023; Gabouer & Bortfeld, 2021). Many classic studies of joint attention have relied on behavioral coding of visual attention. New technology, specifically the rise of eye-tracking, has helped to better objectively operationalize and assess the details of key visual interactions like joint attention (Bradley et al., 2023). Eye-tracking provides a quantifiable metric of visual attention, allowing researchers to gain insight into what the participant is seeing and how they are focusing. Traditional eye-tracking methodology requires the participant to remain stationary while viewing stimuli presented on a screen. This approach has been used particularly with non-verbal populations, including human infants. In developmental research, eye-tracking has been used to examine how, for example, infants' social attention relates to language learning (Tenenbaum et al., 2015) and how toddlers reason about causality (Sobel & Kirkham, 2006).

This approach has also been used to study joint attention through gaze following paradigms. In these, the participant watches footage of a person looking at them and then looking at an item on the screen. Both participant and virtual experimenter then jointly attend to this item (Navab et al., 2012). This kind of responsive joint attention requires that the participant monitors the attention state of the experimenter, following the experimenter's gaze as their attention shifts (Carpenter et al., 1998).

Screen-based eye-tracking has not been limited to children. Non-human primates like chimpanzees have been tested using this approach. Using creative attraction points to keep participants stationary (typically a mounted straw that dispenses juice), researchers have improved our understanding of how chimpanzees examine pictures. We now know that chimpanzees focus on the body and focal figure rather than the background in a similar way to human beings (Kano & Tomonaga, 2009). Screen-based eye-tracking in chimpanzees has also been used to help study complex theory of mind problems, showing that they may pass implicit false belief tasks (Kano et al., 2017).

Screen-based eye-tracking has also been extended to dogs. We know that dogs have worse visual acuity and less sensitive color perception than humans. In contrast, dogs are more sensitive

to flicker rates, and they surpass human visual performance in dim lighting conditions (Byosiere et al., 2018). Researchers have made advances in understanding how dogs respond to visual stimuli using stationary screen-based eye-tracking (Karl et al., 2020a; Somppi et al., 2012). By presenting precise stimuli on screens and recording dogs' eye movements, we know that dogs can recognize photos of familiar humans and conspecifics (Somppi et al., 2014) and that they respond differentially to human faces expressing different emotions (Karl et al., 2020b; Kis et al., 2017; Somppi et al., 2016). Dogs tend to direct their visual attention to living creatures in the foreground (versus the background), a pattern also observed in chimpanzees and humans (Kano & Tomonaga, 2013; Törnqvist et al., 2020).

The human-dog bond has been explored in screen-based eye-tracking systems, evaluating dogs' reactions to modified (i.e., inverted) human and dog faces (Racca et al., 2010; Somppi et al., 2014). Screen-based stationary eye-tracking systems have also helped researchers determine how dogs look at human social-communicative gestures like pointing (Ogura et al., 2020) as well as how they view images of natural scenes (Törnqvist et al., 2020). Although promising, screen-based eye-tracking limits the information that can be presented to video displays and 2D images and may reduce the real-world validity of the results. Further, the requirement for dogs to remain stationary has limited the specific research questions that could be explored, as naturalistic interactions like walking and playing are impossible.

Researchers have begun using mobile, head-mounted eye-trackers to achieve more naturalistic results. Head-mounted eye-trackers capture the first-person view of the wearer as well as record their eye movements, which are mapped together after data collection. They can be worn in various ways (i.e., caps, glasses, goggles), allowing for their use in exploring natural behaviors with head shapes ranging from peacocks to lemurs (Shepherd & Platt, 2006; Yorzinski et al., 2013). Researchers have explored human-chimpanzee interactions using an adapted goggle setup (Kano & Tomonaga, 2013). This method's largest take-up, however, has come from research on young infants. Head-mounted eye-trackers have given us insight into how infants respond to their mother's voices (Franchak et al., 2011) and the differences in how crawling and

walking toddlers see the world (Kretch et al., 2014).

Critically, head-mounted eye-tracking has advanced our understanding of joint attention through development. In head-mounted eye-tracking studies, joint attention is typically operationalized as instances where both participants looked at the same item at the same time. From studies with head-mounted eye-tracking, we know how joint attention can facilitate learning and cooperation. For example, increased instances of joint attention during storybook reading facilitated children's learning of printed words (Guo & Feng, 2013). We also know that joint attention is related to hand-eye coordination (Yu & Smith, 2013) and action prediction (Monroy et al., 2020).

Most recently, head-mounted eye-tracking has been extended to dogs (Pelgrim et al., 2022, 2025). We are now able to use head-mounted eye-tracking to study dogs' visual attention during dynamic or mobile tasks with a similar accuracy to that seen with human infants.

1.2 The Role of Pointing

A major way we interact visually with each other is through the use of social-communicative gestures. One of the most widely studied gestures is pointing. Point following has been studied extensively in children across development starting in infancy. Human beings are generally highly accurate at following points and can identify the target of a point from an early age (Bertenthal et al., 2014; Morissette et al., 1995; Tomasello et al., 2007). At 12 months, we understand that pointing is object-oriented, meaning that we point to indicate a referent. Classic tasks of pointing comprehension in infancy use looking-time based approaches and consistently demonstrate that young children attribute intentions to pointing gestures and predict behavior as a result (Behne et al., 2012; Woodward & Guajardo, 2002).

Infants also begin producing points in the first year of life. We can point flexibly from a young age, using our gestures before we develop spoken language to interact with social partners. Even after developing language, young children regularly use pointing gestures to

serve a variety of purposes. Pointing can be categorized, from this perspective, by the goal or intention of the pointer, and the function the gesture serves. First, pointing can be declarative (or proto-declarative in the absence of language). Declarative pointing is a direction to look at something or to share interest. This style is highly social in nature, and facilitates joint attention, directing the attention of the viewer in a bid to share or engage with something in the world. Production of declarative pointing develops early in the second year of life (Bates et al., 1975; Cochet & Vauclair, 2010). For example, a young child may point to a fountain to direct their parents' attention to the splashing water so they can watch it together. Infants can also use declarative pointing to provide information for others. This is a subtype of declarative pointing, called informative pointing. In informative pointing, infants point to help adults locate lost items, sharing the information they have to assist a social partner (Liszkowski et al., 2006; Tomasello et al., 2007).

Second, pointing can be a social tool, facilitating the accomplishment of goals or the acquisition of items that the infant cannot achieve on their own. Imperative (or proto-imperative) pointing serves to make a gestural request or command to the social partner (Bates et al., 1975; Rohlfing et al., 2022; Tomasello et al., 2007). For example, a young child may point to a toy intending that their parent bring the toy to them. Like declarative pointing, imperative pointing arises at about one year of age (Bates et al., 1975; Carpenter et al., 1998). Imperative pointing can also serve a cooperative goal, requesting assistance when working together (Tomasello et al., 2007).

More recently, pointing in young children has been described as a request for information. Interrogative (or proto-interrogative) pointing serves as an epistemic request by the infant to their social partner (Begus & Southgate, 2012; Kovács et al., 2014; Southgate et al., 2007). Seen in one-year-old children, interrogative points help the infant to acquire new information, rather than trying to provide directions (as in imperative) or share in an experience (as in declarative). For example, a young child may point at an unfamiliar animal at a zoo, intending that their parent tell them about the animal, providing information about the animal's name or what sounds

it makes.

We are able to produce points flexibly to support our goals of engaging with others, learning about the world, and acquiring wants and needs. Similarly, we are able to understand and appropriately respond to pointing gestures in a variety of contexts. Our success at point comprehension is impressive, but it may not be exceptional. It is also a well-established finding that another species, dogs, follow human pointing gestures. From puppyhood and with no explicit training, dogs can follow human pointers to find hidden food items (Agnetta et al., 2000; Hare et al., 2002; Miklósi & Soproni, 2006). Dogs can interpret pointing gestures proto-imperatively, going to a location indicated by the pointer. This ability sets dogs apart from our closest genetic relatives, chimpanzees, who fail to follow imperative points (Kirchhofer et al., 2012). Imperative points are all around dogs, and are frequently used by their owners to give instructions. For example, dog owners may point to say "go upstairs". Dogs can also interpret points in a proto-declarative fashion. For example, dog owners may point to direct their dogs' attention to a treat the dog missed, providing them with information about a lost item (see Bensky et al., 2013, for a review).

Dogs follow human pointing gestures selectively, tending to follow the points of individuals who were previously accurate, knowledgeable, and prosocial (Maginnity & Grace, 2014; Pelgrim et al., 2021; Silver et al., 2020). Their selective following, along with their ability to make use of pointing gestures across tasks, has been suggested as evidence that dogs understand some portion of the communicative intent behind points. How the point is being interpreted, though, is still an open question.

Success at pointing tasks is a common example of visual and cognitive engagement, namely that the producer of the point is engaged in a collaborative action sequence with the point viewer (Tomasello, 1998; Tomasello et al., 2005). Under the shared intentionality hypothesis, collaborative engagement should be unique to human beings. However, non-human animals can follow pointing gestures, as cited above. To resolve this apparent conflict, recent work has aimed to subdivide point comprehension based on the point following task itself rather than

the function underlying the point (Lyn & Christopher, 2020; Tomasello et al., 2007). In this perspective, there are three categories of pointing based on the proximity of the search items and of the tip of the pointing hand. These can be examined in the context of a classic pointing task, where an individual points at one of two items.

Proximal-Proximal pointing occurs when the items are close together, and the point is within a few inches of the correct item. Distal-Distal pointing is where the items are greater than three feet apart, and the point is more than a foot away from the item. In both of these instances, point following can be explained by associative learning mechanisms, namely stimulus enhancement or associative learning (Miklósi & Soproni, 2006; Osborne & Mulcahy, 2019; Povinelli et al., 1999). The third pointing type, Proximal-Distal, involves items that are closer together than in Distal-Distal points, and the point is greater than 12 inches from the target. In the view of Lyn and Christopher (2020), this pointing type is only solvable through triadic engagement and the full suite of unique cognitive abilities that come along with it.

It is appealing to subdivide points to help explain the conflict in cross-species comparisons. This separation is simple and requires no revisions to existing theories. Unfortunately, these distinctions are based on relatively arbitrary cut-offs and rely on broad descriptions of relatively complex gestures. Qualitative descriptions are the norm in describing pointing gestures, and there remains a need to better understand and more precisely describe the gesture itself. Describing all pointing tasks as falling into one of three categories based on the proximity of the pointer and the search items fails to account for other nuances in the tasks themselves. For example, points are often described as being across the body and to a distant target. But how is this gesture actually executed? How is the body of the experimenter moving to produce this gesture? Further, we know fairly little about how non-human animals' visual systems. Understanding how points and other social interactions are seen required building an understanding of the base level of visual attention displayed by this other species. Quantifying pointing gestures and expanding on classic point following tasks will help to situate our understanding of point following in the framework of social communication more broadly.

1.3 The Research Program

In this dissertation, I aim to evaluate a key theory of human unique cognition - the shared intentionality hypothesis. In service of this theoretical motivation, I present a series of empirical studies exploring how humans and dogs visually interact with each other.

In Chapter 2, I explore how dogs examine their visual environment in an ecologically valid task. This provides an important baseline understanding of dogs' visual context, and confirms that we now have an appropriate method to look for key predictions made by the shared intentionality hypothesis about uniquely human visual behaviors. I also demonstrate that dogs' visual attention is not well modeled by a strictly bottom-up low-level visual saliency model, but that they actively direct their visual attention to categories of items differently. This foundation allows me to proceed with examining a unique instance of visual social interaction - point following.

Point following is often referenced as an example of collaborative engagement, and we would expect human toddlers to outperform dogs and follow points at near ceiling levels. In Chapter 3, I compare how dogs and toddlers perform on a complex point-following task. In addition to adding complexity to the task, I incorporate a quantifiable approach to examining pointing gestures that helps to capture precisely how the gesture is being produced. I also demonstrate with an adult sample that the task is easily solvable.

In Chapter 4, I examine how dog owners and their dogs interact on a pointing task. I focus on this naturalistic interaction and quantify how dog owners point for their own dogs. I focus on exploring how dog owners potentially integrate various body language cues to support their dogs in resolving the inherent ambiguity involved in pointing. I also capture how dogs physically approach the task and compare approach and choice patterns to the behaviors observed in Chapter 3.

In Chapter 5, I explore how dog owners work with their dogs on an ecologically valid searching task using a dual eye-tracking set-up. I specifically test for a key marker of "human-

unique" visual interactions, joint attention. If joint attention is unique to human beings, as predicted by the shared intentionality hypothesis, we should not see evidence of this behavior in human-dog dyads. I also explore if joint attention and another common visual behavior, mutual gaze, relate to search task performance.

Finally, in Chapter 6, I summarize the dissertation and discuss avenues for future research. I review key implications from my findings and discuss limitations of the current work.

Chapter 2 is a minimally expanded adaptation of my first-author publication in Cognitive Science (Pelgrim et al., 2025). Chapters 3 and 4 both expand on my first author publication in the Proceedings of the 46th Annual Meeting of the Cognitive Science Society (Pelgrim et al., 2024). Chapter 3 uses a similar task structure and the same analysis approach as Pelgrim et al. (2024). Chapter 4 expands on Pelgrim et al. (2024) using the same procedure, but with a significantly larger sample and novel analyses.

References

Agnetta, B., Hare, B., & Tomasello, M. (2000). Cues to food location that domestic dogs (*Canis familiaris*) of different ages do and do not use. *Animal Cognition*, *3*(2), 107–112. https://doi.org/10.1007/s100710000070

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In *Joint attention: Its origins and role in development* (pp. 131–158). Lawrence Erlbaum Associates, Inc.

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants Parse Dynamic Action. *Child Development*, *72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Behne, T., Liszkowski, U., Carpenter, M., & Tomasello, M. (2012). Twelve-month-olds' comprehension and production of pointing. *British Journal of Developmental Psychology*, *30*(3), 359–375. https://doi.org/10.1111/j.2044-835X.2011.02043.x

Bensky, M. K., Gosling, S. D., & Sinn, D. L. (2013, January). Chapter Five - The World from a Dog's Point of View: A Review and Synthesis of Dog Cognition Research. In H. J. Brockmann, T. J. Roper, M. Naguib, J. C. Mitani, L. W. Simmons, & L. Barrett (Eds.), *Advances in the Study of Behavior* (pp. 209–406, Vol. 45). Academic Press. https://doi.org/10.1016/B978-0-12-407186-5.00005-7

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, *50*(8), 2036–2048. https://doi.org/10.1037/a0037152

Bradley, H., Smith, B. A., & Wilson, R. B. (2023). Qualitative and Quantitative Measures of Joint Attention Development in the First Year of Life: A Scoping Review. *Infant and Child Development*, *32*(4), e2422. https://doi.org/10.1002/icd.2422

Byosiere, S.-E., Chouinard, P. A., Howell, T. J., & Bennett, P. C. (2018). What do dogs (*Canis familiaris*) see? A review of vision in dogs and implications for cognition research. *Psychonomic Bulletin & Review*, *25*(5), 1798–1813. https://doi.org/10.3758/s13423-017-1404-7

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint Attention and Imitative Learning in Children, Chimpanzees and Enculturated Chimpanzees. *Social Development*, *4*(3), 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

Catala, A., Mang, B., Wallis, L., & Huber, L. (2017). Dogs demonstrate perspective taking based on geometrical gaze following in a Guesser–Knower task. *Animal Cognition*, *20*(4), 581–589. https://doi.org/10.1007/s10071-017-1082-x

Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development*, *33*(4), 431–441. https://doi.org/10.1016/j.infbeh.2010.04.009

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*(2), 409–412. https://doi.org/10.3758/BF03203962

Corriveau, K. H., Meints, K., & Harris, P. L. (2009). Early tracking of informant accuracy and inaccuracy. *British Journal of Developmental Psychology*, *27*(2), 331–342. https://doi.org/10.1348/026151008X310229

Ellsworth, P. C., & Ludwig, L. M. (1972). Visual Behavior in Social Interaction. *Journal of Communication*, *22*(4), 375–403. https://doi.org/10.1111/j.1460-2466.1972.tb00164.x

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking: Head-Mounted Eye Tracking. *Child Development*, *82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Gabouer, A., & Bortfeld, H. (2021). Revisiting how we operationalize joint attention. *Infant behavior & development*, *63*, 101566. https://doi.org/10.1016/j.infbeh.2021.101566

Germonpré, M., Lázničková-Galetová, M., & Sablin, M. V. (2012). Palaeolithic dog skulls at the Gravettian Předmostí site, the Czech Republic. *Journal of Archaeological Science*, *39*(1), 184–202. https://doi.org/10.1016/j.jas.2011.09.022

Gray, P. B., & Young, S. M. (2011). Human–Pet Dynamics in Cross-Cultural Perspective. *Anthrozoös*, *24*(1), 17–30. https://doi.org/10.2752/175303711X12923300467285

Guo, J., & Feng, G. (2013). How Eye Gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? In Y. I. Nakano, C. Conati, & T. Bader (Eds.). Springer London. https://doi.org/10.1007/978-1-4471-4784-8_2

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The Domestication of Social Cognition in Dogs. *Science*, *298*(5598), 1634. https://doi.org/10.1126/science.1072702

Hayden-Evans, M., Milbourn, B., & Netto, J. (2018). 'Pets provide meaning and purpose': A qualitative study of pet ownership from the perspectives of people diagnosed with

borderline personality disorder. *Advances in Mental Health*, *16*(2), 152–162. https://doi.org/10.1080/18387357.2018.1485508

Kaas, J. H. (1992). Do humans see what monkeys see? *Trends in Neurosciences*, *15*(1), 1–3. https://doi.org/10.1016/0166-2236(92)90336-7

Kaas, J. H., & Balaram, P. (2014). Current research on the organization and function of the visual system in primates. *Eye and Brain*, *6*(sup1), 1–4. https://doi.org/10.2147/EB.S64016

Kano, F., Krupenye, C., Hirata, S., & Call, J. (2017). Eye tracking uncovered great apes' ability to anticipate that other individuals will act according to false beliefs. *Communicative & Integrative Biology*, *10*(2), e1299836. https://doi.org/10.1080/19420889.2017.1299836

Kano, F., & Tomonaga, M. (2009). How chimpanzees look at pictures: A comparative eye-tracking study. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1664), 1949–1955. https://doi.org/10.1098/rspb.2008.1811

Kano, F., & Tomonaga, M. (2013). Head-Mounted Eye Tracking of a Chimpanzee under Naturalistic Conditions. *PLoS ONE*, *8*(3), e59785. https://doi.org/10.1371/journal.pone.0059785

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135–169. https://doi.org/10.1075/is.7.2.04kap

Karl, S., Boch, M., Virányi, Z., Lamm, C., & Huber, L. (2020a). Training pet dogs for eye-tracking and awake fMRI. *Behavior Research Methods*, *52*(2), 838–856. https://doi.org/10.3758/s13428-019-01281-7

Karl, S., Boch, M., Zamansky, A., van der Linden, D., Wagner, I. C., Völter, C. J., Lamm, C., & Huber, L. (2020b). Exploring the dog-human relationship by combining fMRI, eye-tracking and behavioural measures. *Scientific Reports*, *10*(1), 22273. https://doi.org/10.1038/s41598-020-79247-5

Kirchhofer, K. C., Zimmermann, F., Kaminski, J., & Tomasello, M. (2012). Dogs (*Canis familiaris*), but Not Chimpanzees (*Pan troglodytes*), Understand Imperative Pointing. *PLOS ONE*, *7*(2), e30913. https://doi.org/10.1371/journal.pone.0030913

Kis, A., Hernádi, A., Miklósi, B., Kanizsár, O., & Topál, J. (2017). The Way Dogs (*Canis familiaris*) Look at Human Emotional Faces Is Modulated by Oxytocin. An Eye-Tracking Study. *Frontiers in Behavioral Neuroscience*, *11*, 210. https://doi.org/10.3389/fnbeh.2017.00210

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy*, *19*(6), 543–557. https://doi.org/10.1111/infa.12060

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and Walking Infants See the World Differently. *Child Development*, *85*(4), 1503–1518. https://doi.org/10.1111/cdev.12206

Kubinyi, E., Pongrácz, P., & Miklósi, Á. (2009). Dog as a model for studying conspecific and heterospecific social learning. *Journal of Veterinary Behavior*, *4*(1), 31–41. https://doi.org/10.1016/j.jveb.2008.08.009

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12- and 18-Month-Olds Point to Provide Information for Others. *Journal of Cognition and Development*, *7*(2), 173–187. https://doi.org/10.1207/s15327647jcd0702_2

Lyn, H., & Christopher, J. (2020). A point is not a point is not a point: Reinterpreting three basic kinds of point comprehension. *The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)*, 260–263. https://doi.org/10.17617/2.3190925

Maginnity, M. E., & Grace, R. C. (2014). Visual perspective taking by dogs (*Canis familiaris*) in a Guesser-Knower task: Evidence for a canine theory of mind? *Animal Cognition*, *17*(6), 1375–1392. https://doi.org/10.1007/s10071-014-0773-9

Malle, B. F. (2008). The Fundamental Tools, and Possibly Universals, of Human Social Cognition. In *Handbook of Motivation and Cognition Across Cultures* (1st ed., pp. 267–296). Elsevier.

Marshall-Pescini, S., Colombo, E., Passalacqua, C., Merola, I., & Prato-Previde, E. (2013). Gaze alternation in dogs and toddlers in an unsolvable task: Evidence of an audience effect. *Animal Cognition*, *16*(6), 933–943. https://doi.org/10.1007/s10071-013-0627-x

Miklósi, Á., & Soproni, K. (2006). A comparative analysis of animals' understanding of the human pointing gesture. *Animal Cognition*, *9*(2), 81–93. https://doi.org/10.1007/s10071-005-0008-1

Monroy, C., Chen, C.-H., Houston, D., & Yu, C. (2020). Action prediction during real-time parent-infant interactions. *Developmental Science*, *24*, e13042. https://doi.org/https://doi.org/10.1111/desc.13042

Morissette, P., Ricard, M., & Décarie, T. G. (1995). Joint visual attention and pointing in infancy: A longitudinal study of comprehension. *British Journal of Developmental Psychology*, *13*(2), 163–175. https://doi.org/10.1111/j.2044-835X.1995.tb00671.x

Naderi, S., Miklósi, Á., Dóka, A., & Csányi, V. (2001). Co-operative interactions between blind persons and their dogs. *Applied Animal Behaviour Science*, *74*(1), 59–80. https://doi.org/10.1016/S0168-1591(01)00152-6

Navab, A., Gillespie-Lynch, K., Johnson, S. P., Sigman, M., & Hutman, T. (2012). Eye-Tracking as a Measure of Responsiveness to Joint Attention in Infants at Risk for Autism. *Infancy*, *17*(4), 416–431. https://doi.org/10.1111/j.1532-7078.2011.00082.x

Ogura, T., Maki, M., Nagata, S., & Nakamura, S. (2020). Dogs (*Canis familiaris*) Gaze at Our Hands: A Preliminary Eye-Tracker Experiment on Selective Attention in Dogs. *Animals*, *10*(5), 755. https://doi.org/10.3390/ani10050755

Osborne, T., & Mulcahy, N. J. (2019). Reassessing shelter dogs' use of human communicative cues in the standard object-choice task. *PLOS ONE*, *14*(3), e0213166. https://doi.org/10.1371/journal.pone.0213166

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022). Head-mounted mobile eye-tracking in the domestic dog: A new method. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01907-3

Pelgrim, M. H., Espinosa, J., Tecwyn, E. C., Marton, S. M., Johnston, A., & Buchsbaum, D. (2021). What's the point? Domestic dogs' sensitivity to the accuracy of human informants. *Animal Cognition*, *24*(2), 281–297. https://doi.org/10.1007/s10071-021-01493-5

Pelgrim, M. H., Raman, S. S., Serre, T., & Buchsbaum, D. (2025). Evaluating Dogs' Real-World Visual Environment and Attention. *Cognitive Science*, *49*, e70080. https://doi.org/10.1111/cogs.70080

Pelgrim, M. H., He, I. X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024). Find it like a dog: Using Gesture to Improve Object Search. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*. https://escholarship.org/uc/item/0nk6w9fd

Povinelli, D. J., Bierschwale, D. T., & Čech, C. G. (1999). Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. *British Journal of Developmental Psychology*, *17*(1), 37–60. https://doi.org/10.1348/026151099165140

Racca, A., Amadei, E., Ligout, S., Guo, K., Meints, K., & Mills, D. (2010). Discrimination of human and dog faces and inversion responses in domestic dogs (*Canis familiaris*). *Animal Cognition*, *13*(3), 525–533. https://doi.org/10.1007/s10071-009-0303-3

Range, F., Kassis, A., Taborsky, M., Boada, M., & Marshall-Pescini, S. (2019). Wolves and dogs recruit human partners in the cooperative string-pulling task. *Scientific Reports*, *9*(1), 17591. https://doi.org/10.1038/s41598-019-53632-1

Rohlfing, K. J., Lüke, C., Liszkowski, U., Ritterfeld, U., & Grimminger, A. (2022). Developmental Paths of Pointing for Various Motives in Infants with and without Language Delay. *International Journal of Environmental Research and Public Health*, *19*(9), 4982. https://doi.org/10.3390/ijerph19094982

Shepherd, S. V., & Platt, M. L. (2006). Noninvasive telemetric gaze tracking in freely moving socially housed prosimian primates. *Methods*, *38*(3), 185–194. https://doi.org/10.1016/j.ymeth.2005.12.003

Silver, Z. A., Furlong, E. E., Johnston, A. M., & Santos, L. R. (2020). Training differences predict dogs' (*Canis lupus familiaris*) preferences for prosocial others. *Animal Cognition*. https://doi.org/10.1007/s10071-020-01417-9

Sobel, D. M., & Kirkham, N. Z. (2006). Blickets and babies: The development of causal reasoning in toddlers and infants. *Developmental Psychology*, *42*(6), 1103–1115. https://doi.org/10.1037/0012-1649.42.6.1103

Somppi, S., Törnqvist, H., Hänninen, L., Krause, C., & Vainio, O. (2012). Dogs do look at images: Eye tracking in canine cognition research. *Animal Cognition*, *15*(2), 163–174. https://doi.org/10.1007/s10071-011-0442-1

Somppi, S., Törnqvist, H., Hänninen, L., Krause, C. M., & Vainio, O. (2014). How dogs scan familiar and inverted faces: An eye movement study. *Animal Cognition*, *17*(3), 793–803. https://doi.org/10.1007/s10071-013-0713-0

Somppi, S., Törnqvist, H., Kujala, M. V., Hänninen, L., Krause, C. M., & Vainio, O. (2016). Dogs Evaluate Threatening Facial Expressions by Their Biological Validity – Evidence from Gazing Patterns. *PLOS ONE*, *11*(1), e0143047. https://doi.org/10.1371/journal.pone.0143047

Southgate, V., Maanen, C. V., & Csibra, G. (2007). Infant Pointing: Communication to Cooperate or Communication to Learn? *Child Development*, *78*(3), 735–740. https://doi.org/https://doi.org/10.1111/j.1467-8624.2007.01028.x

Stanley, I. H., Conwell, Y., Bowen, C., & Van Orden, K. A. (2014). Pet ownership may attenuate loneliness among older adult primary care patients who live alone. *Aging & Mental Health*, *18*(3), 394–399. https://doi.org/10.1080/13607863.2013.837147

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Thalmann, O., Shapiro, B., Cui, P., Schuenemann, V. J., Sawyer, S. K., Greenfield, D. L., Germonpré, M. B., Sablin, M. V., López-Giráldez, F., Domingo-Roura, X., Napierala, H., Uerpmann, H.-P., Loponte, D. M., Acosta, A. A., Giemsch, L., Schmitz, R. W., Worthington, B., Buikstra, J. E., Druzhkova, A., … Wayne, R. K. (2013). Complete mitochondrial genomes of ancient canids suggest a European origin of domestic dogs. *Science*, *342*(6160), 871–874. https://doi.org/10.1126/science.1243650

Tomasello, M. (1998). Reference: Intending that others jointly attend. *Pragmatics & Cognition*, *6*(1-2), 229–243. https://doi.org/10.1075/pc.6.1-2.12tom

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Törnqvist, H., Somppi, S., Kujala, M. V., & Vainio, O. (2020). Observing animals and humans: Dogs target their gaze to the biological information in natural scenes. *PeerJ*, *8*, e10341. https://doi.org/10.7717/peerj.10341

Virányi, Z., Gácsi, M., Kubinyi, E., Topál, J., Belényi, B., Ujfalussy, D., & Miklósi, Á. (2008). Comprehension of human pointing gestures in young human-reared wolves (*Canis lupus*) and dogs (*Canis familiaris*). *Animal Cognition*, *11*(3), 373–387. https://doi.org/10.1007/s10071-007-0127-y

Wan, M., Kubinyi, E., Miklósi, Á., & Champagne, F. (2009). A cross-cultural comparison of reports by German Shepherd owners in Hungary and the United States of America. *Applied Animal Behaviour Science*, *121*(3), 206–213. https://doi.org/10.1016/j.applanim.2009.09.015

Woodward, A. L., & Guajardo, J. J. (2002). Infants' understanding of the point gesture as an object-directed action. *Cognitive Development*, *17*(1), 1061–1084. https://doi.org/10.1016/S0885-2014(02)00074-6

Yorzinski, J. L., Patricelli, G. L., Babcock, J. S., Pearson, J. M., & Platt, M. L. (2013). Through their eyes: Selective attention in peahens during courtship. *Journal of Experimental Biology*, *216*(16), 3035–3046. https://doi.org/10.1242/jeb.087338

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

CHAPTER 2

Evaluating Dogs' Real World Visual Environment and Attention [1]

2.1 Background

What does your dog see on a walk? What catches their visual attention? Studying how individuals visually interact with the world from their own perspective gives insight into their cognition in contexts ranging from a pet dog scanning for squirrels to an urban search and rescue dog navigating rubble to find missing people. Egocentric vision research captures the visual information available to an individual, such as the items present in their environment, from their first-person perspective, as well as how they allocate their attention to those items. This approach has become widely used to explore how human infants' visual environments and interactions change over the course of development and how these changes impact, among other things, their developing ability to recognize faces and their acquisition of language. At present, however, real-world egocentric vision research and a better understanding of how they allocate attention to and learn from their real-world environment have been, with some exceptions, mostly limited to humans (Rodin et al., 2021; Smith et al., 2015)

[1]This chapter is minimally adapted from Pelgrim et al., (2024) published in Cognitive Science

The present research uses head-mounted eye-tracking to explore the visual statistics of the environment and gaze behaviors of the domestic dog, a species closely related to humans by emotional bond and relied upon by humans in a variety of working roles and as companions. There is a growing interest in how dogs visually interact with their world. Dogs are relied upon to navigate the human-built visual world in a variety of work settings (i.e., as guide dogs for the blind) and as pet dogs, yet little is known about how they allocate their visual attention to complete these tasks.

It is not currently known how dogs (or canids in general) visually interact with their daily real-world environment, including the kinds of items that are present in their field of view and how they allocate their visual attention. Dogs, like many species, might perceive their world as structured into high level visual categories. In particular, dogs may use a primarily top-down approach, actively directing their visual attention to certain categories or classes of items. In this case, dogs, like human infants by approximately six months of age (Hunnius & Bekkering, 2010; Reynolds, 2015), could be visually attending to objects in a sophisticated manner. Dogs likely have different mental classification of items than human beings, but they may primarily be responding to the type or category of object. Dogs may also display signs of anticipatory looking, directing their attention to locations where items of interest could appear. On the other hand, it is also possible that dogs may only have our shared bottom-up processes but lack any categorical or identity based understanding of their visual world. This option goes beyond dogs having a different categorical representation of the world than human beings. Dogs' vision could be best explained by a bottom-up process, where they are primarily responding to low level visual cues and saliency. If this is the case, we would expect that a model of image saliency, a bottom-up approach that identifies the most salient region of an image on the basis of low-level cues like brightness and luminance, would capture dogs gazing behaviors. Finally, while unlikely, as a possibly predominantly olfactory species, dogs' visual attention may be random or driven by their noses and not by visual features of the environment. If true, we would expect gaze to be randomly distributed.

We aimed to capture how dogs direct their visual attention using head-mounted eye-tracking to record what items were in the dogs' field of view (their relative frequencies and sizes) and which of the items they looked at. We also incorporated computer vision techniques to identify these items, something that was previously extremely labor intensive. Finally, we compared dogs' visual attention behaviors to existing image-saliency models to explore if dogs' gazing behaviors were accurately captured by low level visual-perceptual cues. Identifying how dogs direct their visual attention outdoors is particularly relevant given the numerous real-world working roles dogs serve in, where they must traverse complex outdoor environments.

2.1.1 Head-Mounted Eye-Tracking

Describing what participants have visual access to is inherently useful and necessary to explain behavior and available cognitive mechanisms. Head-mounted eye trackers have given us insight into how infants respond to their mother's voices (Franchak et al., 2011) and how infants and parents coordinate joint attention to items (Yu & Smith, 2013). They have also been used to capture how visual attention changes over development. As children transition from crawling to walking, they move from looking at the floor in front of them while in motion to looking at walls, items, and caregivers. This changes their frequency of looks to caregivers because in order to look at their caregiver, crawlers have to stop and either sit or crane their heads, whereas walkers can stay in motion and look ahead (Kretch et al., 2014). An improved understanding of infants' visual experiences has informed theories on language acquisition, namely that a statistical learning framework can reasonably be applied to word learning because of the distribution of item frequencies in naturalistic scenes. For example, the first nouns that children learn tend to be those items most frequently present in their environments (e.g., spoon, bowl) (Clerkin et al., 2017; Smith et al., 2018). In sum, the items present in a child's environment support their learning and development. Capturing the types and distribution of items in children's visual environments has informed our understanding of their development across cognitive and social domains (Jayaraman & Smith, 2020).

The present study seeks to describe and categorize both the visual information available to dogs in their daily environments, as well as characterize how they direct their attention within that space. We took an ecologically valid approach to understanding dogs' attention in their daily environment by having dogs walk with their guardians in a normal fashion along a predetermined route.

Our study had three major aims. First, we explored the relative frequency that items were present in dogs' field of view, providing us with an understanding of the items available for dogs to look at in an ecologically relevant context. Second, we evaluated how dogs looked at the items in their field of view, exploring for consistency in looks across exposures to particular item classes (e.g., people) when those items were present in the environment (i.e., if they looked at a person each time there was a person in their field of view or if they rarely looked at people while people were in their field of view). We evaluated this using the proportion of time dogs fixated on the item relative to the amount of time that item was in their field of view. We particularly wanted to explore the social domain, namely if people were looked at consistently across exposure and if they were looked at more than other non-social items. We also examined if dogs' visual attention to items in their field of view was predicted by the item categories present in their visual environment or could instead be explained by low-level properties of items rather than the identity of the item class, using image saliency analysis. Our third aim was to look for any individual differences in visual attention to item classes between dogs, considering both the items in their view and which of those items they looked at. It is possible that dogs may differ in what items they find visually interesting, which could result in differences both in the items in their field of view and in which of those items they looked at.

Historically, this work would have been logistically challenging due to the time required to annotate the items present in the dogs' field of view. We developed a computer vision pipeline to identify each of the items in the image on a pixel level, segmenting out the individual items from the broader image and then applying class labels to each pixel-identified item. Potential classes were determined in advance based on common items present across multiple dogs' walks. We

then integrated the eye-tracking data to provide a label for the item the dog was gazing at for each fixation.

2.2 Behavioral Methods

2.2.1 Ethics Note

This study was approved by the Brown University Institutional Animal Care and Use Committee (IACUC), protocol number 20-05-0002. Procedures were in accordance with the ASAB/ABS Guidelines for the Use of Animals in Research and complied with the United States Department of Agriculture Animal Welfare Act & Regulations. All participants were pet dogs who were walked by their guardians throughout the session.

2.2.2 Participants

Participants were 11 dogs (Female = 7, Mean Age = 6.5 Years) recruited for participation in a broader eye-tracking training program. Dog breeds represented were 1 Miniature American Shepherd, 1 Australian Labradoodle, 2 Golden Retrievers, 2 Labrador Retrievers, and 5 Mixed Breeds. Each dog was recorded walking along the route once. An additional 2 dogs were excluded from data analysis due to camera displacement (n = 1) and a refusal to walk while in the goggles (n = 1). Dogs were chosen for suitability with the eye-tracking training program and the guardian's willingness to complete the training. Prior to participation in this experiment, dogs were trained at home by their guardians to wear the eye-tracking goggles using commercially available dog goggles following the methods described by Pelgrim et al. (2022). The purpose of this training was to acclimate dogs to the goggles so that they were comfortable wearing them. Dogs were approved to begin participation if guardians reported they were comfortable walking and behaving normally at home and outdoors, wearing training goggles for at least 10 minutes.

2.2.3 Procedure & Materials

Throughout sessions, dogs wore a custom-developed head-mounted eye-tracker consisting of two cameras affixed to dog goggles (Positive Science, Inc.). One camera records the dog's right eye via an infrared eye camera with an adjacent infrared emitting diode (hereafter the eye camera). The other camera (hereafter the scene camera) recorded the dogs' first-person perspective, recording a field of view of 101.55º horizontal and 73.60º vertical. Videos from both cameras were digitized at 29.96 frames per second. This apparatus, acclimation, and calibration procedures are adapted from comparable models in other species and has been validated in dogs using alternative methods (Figure 2.1; Pelgrim et al., 2022). Dogs also wore a harness throughout their session to hold the video recording pack and its battery.

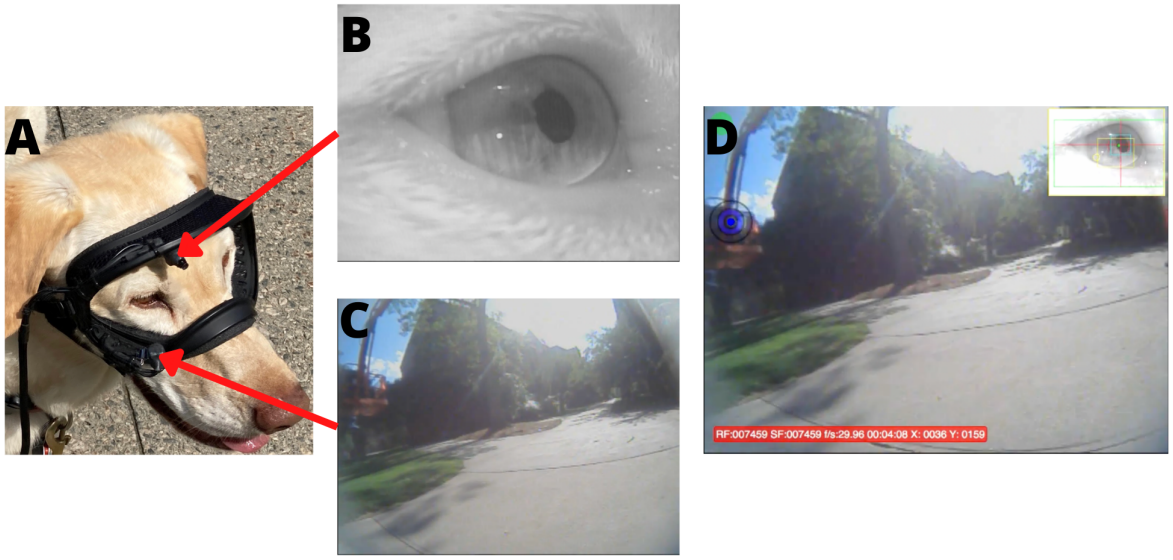

Figure 2.1: (A) A dog wearing the eye-tracking goggles which had two cameras. (B) The dog's eye recorded from the eye camera. (C) The dog's view recorded by the scene camera. (D) The dog's view with their gaze indicated by the blue circle.

Prior to starting their walk, dogs first completed a calibration procedure. This procedure allowed for the dogs' eye movements as recorded from the eye camera to be mapped onto

the field of view recording from the scene camera, the result being the dogs' gaze (or where in their environment they were looking), which could be extrapolated offline for the entire recording, after the session (Pelgrim et al., 2022). Consistent with previous work, we calibrated the eye-tracker when the experimenter drew dogs' attention to specific points using a treat. The dog's guardian held their dog's head stationary while the dog followed treats held by an experimenter, via eye movements alone, through 5 unique points in space. Each point in space where the dog looked at the experimenter provided a known point, meaning that the positioning of the dogs' pupil and corneal reflection was linked to where in their first-person view they were looking. The 5 points were chosen to be spread widely across the dogs' first-person view, thus requiring a wide range of eye movements. These eye movements made the offline extrapolation of the point of regard for the entirety of the walk using eye-tracking software more accurate. During the calibration procedure, a removable handheld screen was plugged into the recording pack to allow the experimenter to verify the eye camera was recording a clear and centered image. After the dog looked at the experimenter in all 5 points, as judged by the experimenter, the LCD screen was removed. The calibration procedure was completed both before and after the walk to provide enough known points for 1) a successful calibration and 2) verification of that calibration accuracy (more details in Data Coding & Preparation).

Following the first calibration, dogs walked with their guardians, following the experimenter's directions, along a pre-set route. The route was 0.5 miles and was chosen for its variety of scenery, including both city streets and quiet campus greenspaces. Guardians were instructed to walk their dog as normal, and the experimenter followed behind or adjacent to the dog-guardian pair. If at any point during the session, the eye-tracker was disturbed or shifted (i.e., the dog shook their body or brushed against a wall), the fit was adjusted, and the eye image was verified. In the event of tracking disruption, the calibration procedure was also repeated.

2.2.4 Data Coding & Preparation

After the session, video data recorded from the eye and scene cameras were combined as described above, using between 4 and 9 calibration points (Yarbus eye-tracking software, Positive Science Inc.). This process identified both the timing and the direction of dogs' gaze, referred to from here on as fixations. Fixations were defined as a stable eye positioning lasting for 100 ms or more as determined by the eye-tracking software in keeping with past work (Pelgrim et al., 2022). More specifically, the timing (start and stop in ms) and the dog's point of regard in the visual scene (defined by x-y coordinates from the scene camera) was identified for each fixation.

This eye-tracking system has previously been established to have a spatial accuracy of approximately 2 to 4º in humans (Franchak et al., 2011; Watalingam et al., 2017), around 4º in peacocks (Yorzinski et al., 2013) and around 3.6º in dogs indoors (Pelgrim et al., 2022). The spatial accuracy of the eye camera mapping onto the scene for our outdoor walks was calculated, as in past work, using the unused points from the calibration procedure. The distance between the extrapolated point of regard (where the eye-tracking software calculated the dog to be looking) and the known point of regard (where the dog was known to actually be looking, namely at the treat bag in the experimenter's hand) was calculated for approximately 20 frames across the calibration points not used for the mapping, providing the spatial accuracy of the mapping, in degrees, for each dog.

The spatial accuracy for the present sample was 5.32º. This is less precise than past implementations mentioned above. However, it was to be expected given that this is the first time this system has been deployed outdoors under natural variable lighting conditions. Further, the spatial accuracy of the tracking for each dog was accounted for when determining dogs' fixated items, creating a region of fixation for each look.

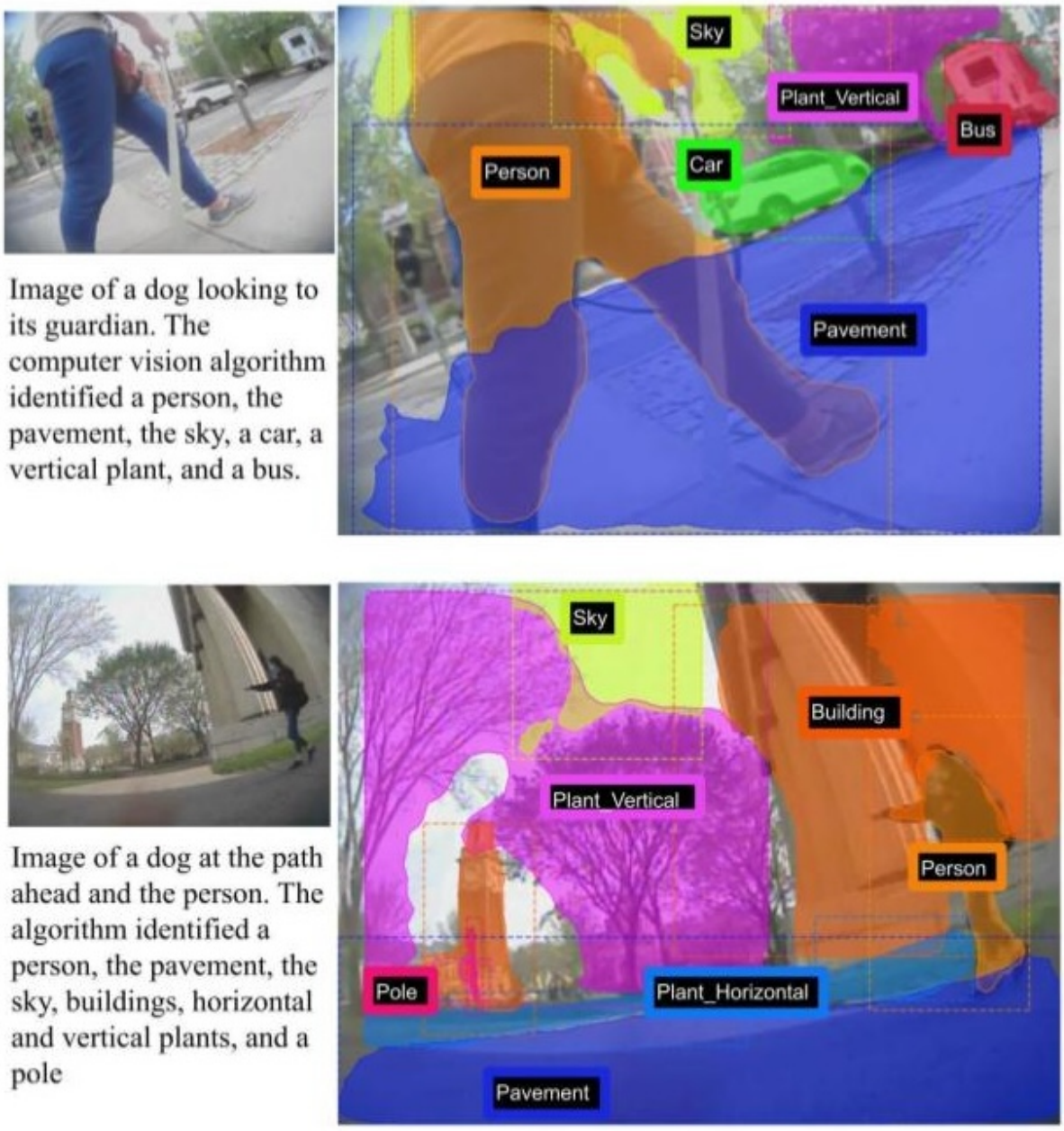

Figure 2.2: Two examples of images segmented from items from the test set, using our MaskR-CNN approach. **Left** - original image, captured by the head-mounted camera. **Right** - fine-tuned MaskRCNN's predicted segmentation masks. Shaded field indicates the identified item and the dashed rectangle shows the outer boundaries of each item. Unique class labels are bolded for easier visualization.

Each fixation was segmented into the items present (and therefore available to be looked at), and coded for the classes of those items (see Figure 2.2), the identity of the item (or items) the dog was looking at, and the relative proportion of the dog's field of view taken up by each

item (a function of both the item's proximity to the dog and the item's absolute size, see below), using computer vision techniques (see Computer Vision section). One limit of our relative size calculations is that our eye-tracking cameras do not record depth information, so object size and proximity cannot be separated (objects that are close will be recorded as larger, large objects in the distance will be recorded as smaller). The classes of items were chosen ahead of data collection as they were consistently present on the pre-determined route. 15 items were chosen in total based on the items available in the real world for multiple dogs (Table 2.1).

The primary focus of this study was how dogs interact with their world visually, so instances where dogs were relying on another sense (such as while sniffing), were removed. Sniffing bouts were defined as looks where two or fewer items were present in the environment. As an example, when a dog was sniffing a pole, the only items visible in the environment were the pole and the plant and both were so close to the camera that it is unlikely that the dog was visually considering them.

In some cases, the dog could be fixated on at a location occupied by more than one item (see Figure 2.3). To determine which item(s) the dog was gazing at, we examined the items identified by our computer vision output that were present in the fixated region (the target of fixation plus the calibration margin of error for each dog). For each item present in the fixated region, we weighted the item according to how much of the fixated region it occupied (Figure 2.3). For example, as in Figure 2.3, if approximately 50% of the pixels in the fixated region were occupied by an item in the class horizontal plants, and approximately 25% each by items in the classes vertical plants and buildings, the fixation weight would be 0.25 each for the building and vertical plants, and 0.5 for the horizontal plants. For each fixation, we also weighted the duration of time spent looking to each class contained within the fixation region based on the proportion of the fixated region that item occupied (i.e., a 200ms fixation to a region containing 50% sky and 50% building would count as 100ms to each of the two classes).

Table 2.1: Looking Behaviors to Classes Across Dogs

Object	**Avg. proportion of time in view (SE)**	**Avg. proportion of time fixated while in view (SE)**	**Avg. size when in view (SE)**
bench / chair	0.033 (0.005)	0.012 (0.004)	0.01 (0.001)
bicycle	0.024 (0.007)	0.099 (0.032)	0.031 (0.003)
building	0.878 (0.022)	0.144 (0.012)	0.145 (0.015)
bus	0.008 (0.005)	0.348 (0.094)	0.187 (0.024)
car	0.299 (0.05)	0.064 (0.015)	0.027 (0.003)
construction	0.011 (0.003)	0.145 (0.04)	0.044 (0.007)
pavement	0.885 (0.047)	0.381 (0.069)	0.336 (0.027)
person	0.389 (0.067)	0.157 (0.034)	0.131 (0.017)
plant_horizontal	0.616 (0.041)	0.174 (0.032)	0.18 (0.019)
plant_vertical	0.934 (0.018)	0.269 (0.058)	0.212 (0.02)
pole	0.168 (0.019)	0.027 (0.008)	0.022 (0.003)
scooter	0.008 (0.002)	0.077 (NA)	0.012 (0.006)
sculpture	0.037 (0.005)	0.036 (0.01)	0.015 (0.004)
sign	0.013 (0.003)	0.049 (0.011)	0.028 (0.009)
sky	0.838 (0.022)	0.07 (0.026)	0.075 (0.01)

Note. Only one fixation to a scooter occurred, so there is no SE for that class.

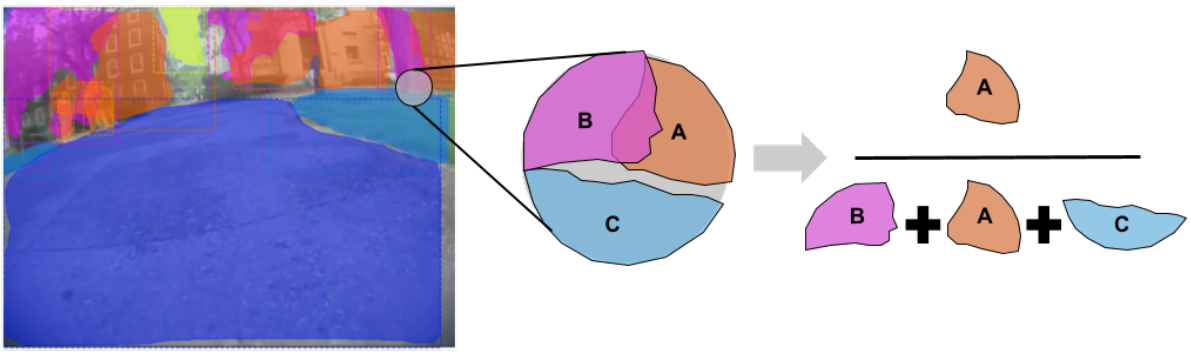

Figure 2.3: The procedure used to estimate the probability distribution over fixated objects. Left-to-right: given the fixation region in the image identified in grey (Left) we consider portions of predicted object masks that intersect the fixation region (Center) in our estimate. We estimate the probability of the fixation towards a given class (e.g. class "A") by normalizing the total number of pixels in the intersection belonging to the given class (i.e. the pixels in intersection belonging to "A") over the total number of pixels belonging to all classes in the intersection (i.e. pixels in intersection belonging to "A", "B" and "C")

2.2.5 Image Saliency Analysis

Dogs could be directing their attention based on interest in different item classes, or it could be a more bottom-up process driven by low-level image features contributing to increased saliency of regions of the image. To examine what a bottom-up visual approach would look like, we conducted image saliency analyses using Saliency Toolbox Version 2.3, a well-established and validated model (Walther & Koch, 2006). For a given image, Saliency Toolbox generates feature maps for standard low-level image features (i.e., luminance, color intensity and opponency, orientations) which are then combined, eventually resulting in a saliency map. Dogs have different color vision than humans do, so to account for this we conducted our image saliency analysis first in greyscale, and then in full color. This allowed us to compare dogs' observed real-world fixations to their expected fixations within the same images if their gaze was primarily driven by low-level image features. If fixations predicted by the saliency model generally align with dogs' real fixations, it would suggest that dogs' visual attention is primarily driven by low-level image features. In contrast, if the saliency model is a poor predictor of dogs' real-world fixations, while item class is a better predictor, it would suggest that dogs' visual attention is being driven by more complex factors, including item category or identity.

2.3 Computer Vision

2.3.1 Methods

Our head-mounted eye-tracking approach generates an average of 1063 fixations in an 11 minute walk. It is therefore not practical to manually annotate the items present in each fixation. Manually annotating the items present from the look image at each fixation would have taken around 1,200 hours based on the time taken to annotate the dataset used for training and testing our computer vision approach. Using computer vision significantly reduced the time required to identify the items in dogs' look images. After the computer vision approach was trained, each dog's complete walk could be annotated in a matter of minutes. Our aim was to automate

the segmentation of images into items, the classification of those items, and the identification of the items the dog was gazing at. To do this, we utilized an automated computer vision method for panoptic segmentation (simultaneous identification of items within the image and their item classes). This approach served to identify and label the item classes present at each fixation (Figure 2.2). We used an off the shelf object identification algorithm (Mask R-CNN with ResNet-101-based Feature Pyramid Network backbone (He et al., 2018) that was pre-trained on Microsoft Common Items in Context, a publicly available image data set (Lin et al., 2015). We then adapted and trained it using a training set of annotated scene images taken from our own data. We used 610 manually annotated images from our dataset for training and an additional 621 images testing. Images were annotated by human coders who were trained to identify each of the item classes. Coders were instructed, for each image, to create a polygon around each of the items from our identified list. Any items not in the class list were left blank. Training and test sets contain comparable distributions of item classes. See Figure 2.2 for an illustration of how the model functions. We find that our fine-tuned model performs well in matching the ground-truth annotations from our test set. After confirming our model's performance on the test set we ran it on our remaining data.

After the images were segmented into items and labeled, we integrated the eye-tracking fixation data (in conjunction with the spatial accuracy of the eye-tracking system) to create a probability distribution of the potential fixated item class, in order to create a fixation weight as described above (Figure 2.3). We used the calibration error of the eye-mounted camera to estimate the radius of error around the fixation pixel; we assumed that the true fixation is expected to lie within this fixation region. To estimate a distribution over fixated items, we normalized the pixel counts for each identified object within the intersection by the total number of pixels in the fixated region assigned to an item class; this generated a probability distribution over possible fixation items for the image. Figure 2.3 highlights a visual example of this probability distribution computation, as discussed earlier.

Across the 11 dogs, a total of 11,698 fixations were recorded spanning 102,868 frames.

After eliminating sniffing bouts, or look images that contained 2 or fewer items in the view as described in data coding and analysis, as well as those that did not result in any predictions from our computer vision analysis, 10,296 look images remained. This set was used in subsequent calculations and analyses.

2.3.2 Computer Vision Results

We first verified that our computer vision model generalized. To be successful, our computer vision algorithm needed to accurately identify item classes across the whole of the image, matching in overall coverage and class label with our human annotated images (our ground truth). To evaluate our model performance, we evaluated performance across metrics, specifically overall image coverage by identified items, class-specific coverage, and fixated region class identification. We also examined both the true positive rate, false positive rate, and the confusion on a class-by-class level (Figure 2.4). All model evaluations were done by comparing our model annotated test set to the ground truth data.

First, we compared the overall coverage of the image from our computer vision model to our ground truth. The average percent of coverage over the frame (consisting of the total area of all model identified item classes) across the 4 dogs in the test set, was smaller in area than ground-truth masks. Our fine-tuned model covers on average 84% of the image, relative to the 93% average of our human annotators. Given this reduction in coverage, we next wanted to evaluate if this resulted in changes in the classes identified in our fixated regions, a critical part of our analysis. We found that they did not significantly impact class distributions for fixation predictions.

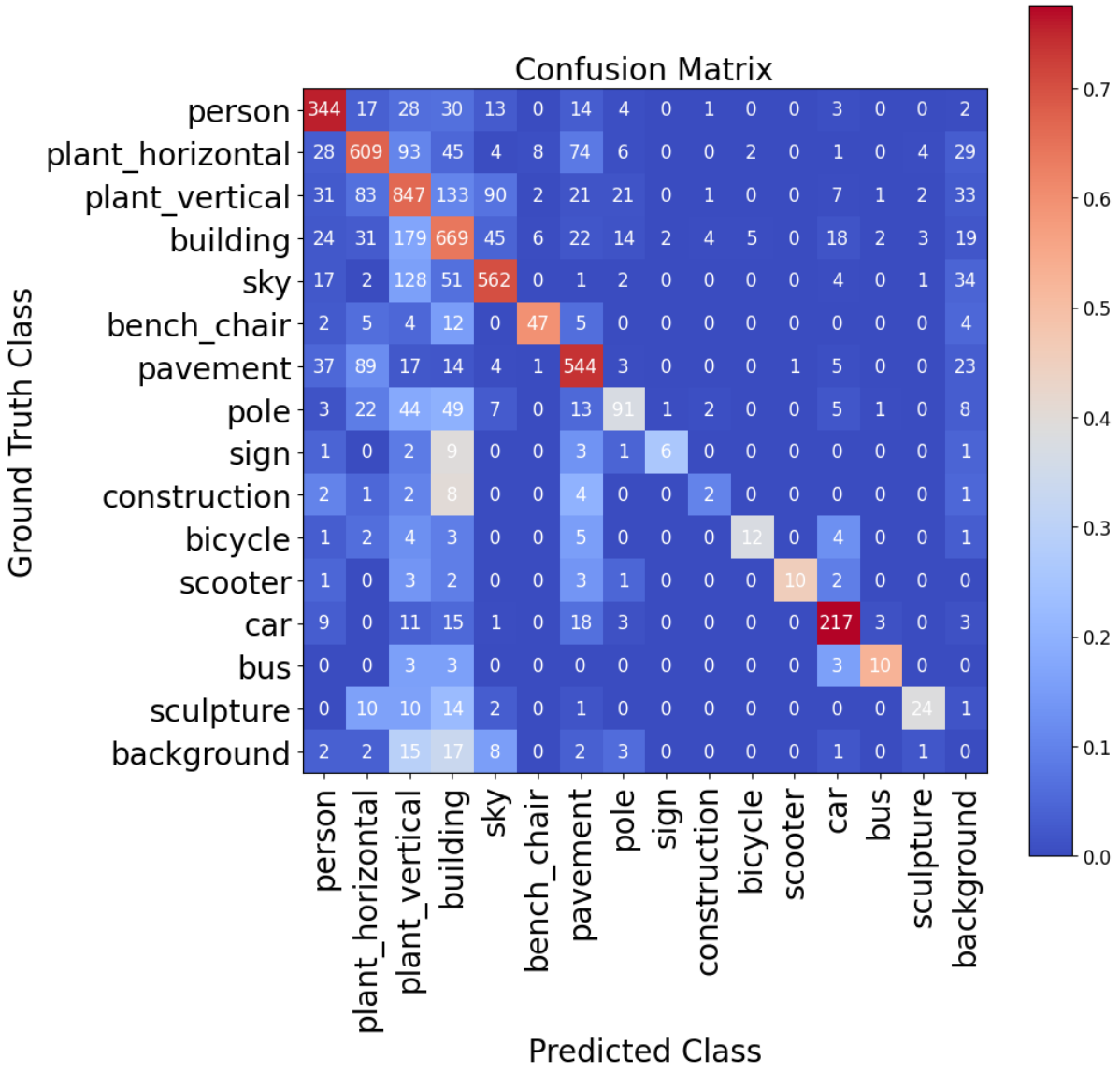

Figure 2.4: Confusion matrix comparing, for our test set, the frequency of identification of an instance of a class relative to its ground-truth occurrence. Items classes as ”background” mean that our algorithm failed to identify the item as a class.

Next, we examined the overlap of the generated classes to the ground-truth classes. We evaluated the Intersection Over Union (IoU) or the number of pixels from each of the model identified classes overlapping with the ground-truth classes, normalized by the number of pixels in the union of the predicted and ground-truth identified classes. We found that the fine-tuned model achieves IoU substantially greater than chance – indicating that the generated masks identifying the item classes are semantically relevant. The median IoU across all classes is 60%

(10 times higher than chance) with a maximum IoU of 85% for the "bus" class and a minimum IoU of 42% for the "pole" class. In comparison, the maximum IoU by random chance is 8% for the "sky" class.

2.4 Statistical Analyses

As a first step to understanding how dogs' view natural scenes, we quantify the items present in dogs' field of view. We report on the proportion of time each item class was present. We then conducted a 15 (item class) by 11 (dog identity) ANOVA to explore the impact of class and dog identity on the average proportion of time items were in view for each dog.

Our next goal was to examine how dogs directed their visual attention at the items present in their environment. We considered a number of factors that could impact how dogs look at item classes, specifically class identity, dog identity, and item size. To explore the impact of dog identity and item type, we conducted a a logistic regression with a firth's correction using the logistf package in R exploring the coded binary fixation data as a function of the item class, dog identity, and an interaction between dog and item class. To conduct this analysis, we used a subset of the classes, exploring only those classes (11/15) that were present in all dogs' views (removing buses, signs, construction equipment, and scooters). We also conducted post-hoc pairwise comparisons using a Tukey correction. To examine the impact of object size, we first conducted a Spearman correlation relating the average size of the item class to the time fixated on that class (relative to the time it was in view). Next, we used a linear mixed effects model to integrate the possible contributors of visual attention, examining the fixation duration as a function of the fixated item class(es), size of fixated item(s), and a possible interaction between the size and item class. We also included a random intercept for each dog.

As a final step, we evaluated the saliency models' predictions for where dogs would look, if their attention was driven primarily by low level image features, against our fixation data by evaluating the Area Under the Curve (AUC) of the Receiver Operating Characteristics (ROC)

(Bylinskii et al., 2019). This compares the accuracy, using true and false positive rates, of the saliency model, relative to the ground truth from the fixation data. We used AUC-Judd, a variant on AUC with a repeated threshold approach. For each potential threshold on a pixel-by-pixel level, all points above the threshold are considered "fixations". Instances where the actual fixation overlaps with the predicted "fixations" based on threshold are considered true positives, and the points that are above threshold but do not overlap with the true fixated region are considered false positives. The ROC curve is plotted from these two values, and AUC-Judd scores can range from .5 (random classification) to 1 (perfect performance). (Bylinskii et al., n.d.; Judd et al., 2012).

2.5 Behavioral Results & Discussion

2.5.1 Items in View

Our first aim was to explore the items present in dogs' field of view during their walks. This is a necessary first step to understand how dogs direct their attention within their field of view. We found a significant effect of item identity, $F(14,140) = 172.07, p < .001, \eta_p^2 = 0.95$, confirming that some item classes were more commonly present in dogs' view, and available to be looked at, than others. On average, dogs tended to have common scene components–plants (horizontal and vertical), pavement, buildings, and the sky–present in their field of view on the majority ($> 50\%$) of their fixations. In addition to these ubiquitous items, dogs had certain categories of items available for gaze moderate amounts of time (< 50 % and > 10 % of fixations). These included people, cars, and poles. Finally, some categories were rarely in dogs' view, including sculptures, benches/chairs, bicycles, signs, construction equipment, buses, and scooters. See Table 1 for a summary of the proportion of time all coded item categories were in view. We also found no effect of dog identity, $F(10,140) = .879, p = .55, \eta_p^2 = 0.06$, meaning that the dogs in our study encountered similar item classes along their walk, and did not differ in the proportion of time that different item categories were in their view. Given that dogs walked

the same route, they had the opportunity to orient their fields of view towards many of the same static items (e.g., buildings) for comparable amounts of time. However, it is still notable that dogs' different sizes, training experiences and personalities did not result in them directing their heads to markedly different items on their walks.

2.5.2 Image Saliency Analysis

For our full color analyses, AUC-Judd was 0.67. Similarly, the AUC-Judd for greyscale analyses was 0.66. 0.5 is expected by chance, and 0.92 is the maximum (Judd et al., 2012). This suggests that the saliency model is a poor predictor of dogs' looking behaviors, or that it's unlikely dogs' fixations are driven solely by low-level image saliency features.

2.5.3 Relative Time Looking to Items

Our second aim was to explore how dogs allocated their visual attention to the items in their view (Figure 2.5). As mentioned previously, it is possible that dogs are not attending to items on the basis of class, but are looking randomly at the items in their view. In an extreme case, dogs could not be actively directing their visual attention at all, passively encountering classes of items as they appear. If this is the case, we would expect dogs' fixations to items to be uniform across classes, when controlling for the time those items were in view. In contrast, dogs may be actively directing their visual attention, selectively looking to certain classes more than others when they are available in the environment. Dogs may also differ in the way they direct their visual attention, and this individual difference may interact with the class identity. For example, some dogs may fixate on a class the majority of time it's in their view, while others never look to that class.

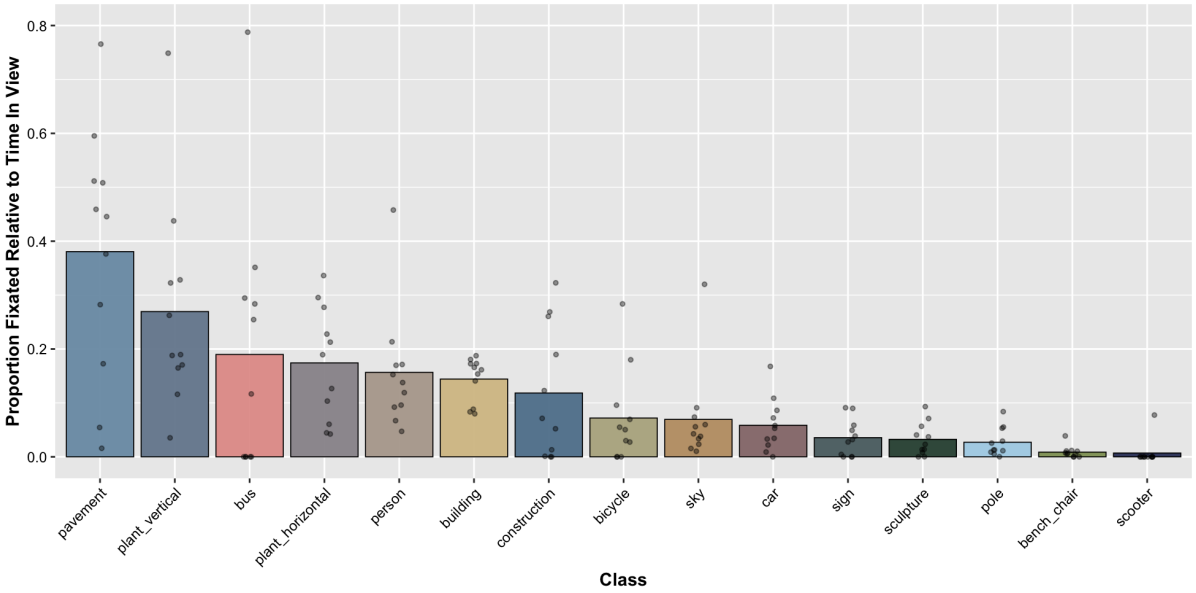

Figure 2.5: Average proportion fixated to each class, relative to the time the class was in view. Dots indicate individual dogs' averages.

We first explored in a binary fashion how often dogs fixated to an item class when it was in view, and if this differed across dogs or potential dog-class interaction. We found a significant main effect of item class, $\chi^2(10) = 1348.97, p < .001$. This suggests that dogs are not uniformly or passively observing the items in their field of view but are actively directing their attention to certain classes of items. For example, dogs looked frequently to people, 15.7% of the time they appeared. In contrast, dogs only looked to benches and chairs around 1.2% of the time they were in view, and post-hoc pairwise comparisons exploring these class differences with a Tukey correction revealed that looks to these two classes were significantly different, $t(51368) = -5.805, p < .001$. This was one of many significant differences by class. Dogs also looked significantly more to people than other asocial classes like sculptures, $t(51368) = 5.137, p < .001$, and the sky, $t(51368) = 16.346, p < .001$.

In past screen-based eye-tracking research in dogs, plants and sky have been used as the background material to explore how dogs look at the primary subject of the image, typically a person or animal present in the foreground (Törnqvist et al., 2020). Interestingly in our sample, dogs were interested in plants, and looked at them relatively frequently. Dogs looked

to vertical plants (like trees and bushes that could not be easily moved through by the dog) 26.9% of the time they were in view. This was significantly more than they looked at people, $t(51368) = -17.154, p < .001$. In contrast, another frequent background item from screen-based eye-tracking, sky, was almost never actually looked at by dogs. The sky was nearly ubiquitous in dogs' view, present and available for view on average 83.8% of fixations, yet dogs looked infrequently at it, only looking at the sky 7% of the time it was in their view (Table 1). This suggests that, unlike plants, in a real-world context, the sky is treated by dogs as background, or at least not something that is worth attending to. Dogs also attend frequently to the ground in front of them, looking to the pavement 38% of the time it's in view and looking to horizontal plants (like grasses that could be walked on) around 17.4% of the time in view. These looks may be for navigational or wayfinding purposes, as past research in adult humans and in children suggests that when navigating in a real-world context, people often look to the ground, helping us to map a path forward and to avoid obstacles. In prior work adults look to the ground 20% to the ground while navigating parking garages de Winter et al. (2021) and 55.9% to street edges and the ground while navigating urban streets Simpson et al. (2019). When navigating obstacles, adults looked to the obstacles on the ground 31.8% of the time, much less than children from the same study who looked at the obstacles 58.9 % of the time Franchak and Adolph (2010).

From the same binary fixation analysis, we found no significant main effect of dog identity, $\chi^2(10) = 4.15, p = 0.94$. However, we did observe a significant interaction between dog identity and class on fixation behaviors, $\chi^2(100) = 3997.19, p < .001$. This suggests that there were individual differences in the priority with which dogs fixated to classes. For some classes, there were significant differences between dogs, and dogs were not all alike in which classes they found visually interesting. As an example, when considering the relative amount of time dogs spent looking to people in their environment, one dog looked at people 46% of the time they were in her view. Post-hoc pairwise comparisons using a Tukey correction found that she looked to people more than most other classes, including items like buildings, $t(51368) = -5.96, p < .001$,

horizontal plants, $t(51368) = 10.843, p < .001$, and sculptures, $t(51368) = 4.60, p < .001$. In contrast, another dog looked at people only around 5% of the time they were in her view, and pairwise comparisons showed she did not look significantly more at people than at any other class ($p > .05$). Post-hoc pairwise comparisons using a Tukey correction also found that these two dogs different in how much they looked at people, $t(51368) = 4.21, p = .001$. In contrast, for other classes like benches and chairs, dogs were quite consistent in their looking behaviors. On average, dogs looked rarely to benches and chairs when in view and there were no significant ($p < .05$) individual differences in fixations to benches and chairs. Our results suggest that dogs differ in how visually interesting they find some, but not all, classes of items. There is little doubt that dogs do experience the world differently, and their size and life experiences may play a role in this.

2.5.4 Impact of Item Size

In addition to the identity of the fixated items present in dogs view, we explored how dogs looked at items as a function of class size. It's possible that dogs are not gazing selectively, but are simply gazing at the things in their environment that take up the most space. If they're selective, we could expect that dogs would occasionally look to items that are comparatively smaller, either because the object itself is smaller or because it's far away.

Across walks, items differed in size, both within and between items. We measured size by the proportion of the dogs' field of view they occupied. Some items, like sculptures and benches/chairs were consistently small in dogs' views ($M_{sculpture} = .01, SD = .04, M_{bench/chair} = .01, SD = .01$). In contrast, other items tended to take up large portions of dogs' view and varied widely in size (i.e., $M_{pavement} = .35, SD = .17, M_{planthorizontal} = .19, SD = .19$). On 4,453 fixations, or 43.25% of fixations they fixated on things that were not the largest item, the largest class in view was one of the fixated items, as determined from the probabilistic fixation data across the region of fixation. On just over half of their total fixations, dogs looked to the largest item class in their view at that moment. On these looks, the object took up an average of

$M = .46, SD = 0.14$ of the dogs' field of view.

The average size of the item was correlated with dogs' fixation to that item. Using a Spearman correlation, we found a significant positive correlation between average size of item class and time fixated on that class (relative to the time it was in view), $r_s = .751, p < .001$. While it is interesting that dogs are tending to look at larger items, conclusions from these results should be limited, as the directionality is unknown. More specifically, it's likely that dogs will turn their heads and/or move toward items of interest, thus making them bigger. However, it's also likely that larger items may be more attention-grabbing, and dogs may be more likely to see them (vs. smaller items which could be equally or more interesting but more easily missed). In this study, we were not able to determine the distance of the items which further limits potential conclusions about the impact of item size on dogs' visual attention.

A major advantage of head-mounted eye-tracking is the use of real-world stimuli and mobile participants, and in this case, our results suggest that dogs may consider at least some plants to be items of interest, exploring them visually and through other sensory modalities (the majority of sniffing bouts that were removed included plants and poles). However, they do not appear to consider the sky as an item class worth investigating. Further research is needed to make more nuanced conclusions about dogs' response to plants and other potential background items.

We explored the duration of each of dogs' fixations ($n = 10,296, M = 301.8$ ms, $SD = 481.09$ ms) as a function of the class of item they were looking at, the size of that item they were looking at, and a possible interaction between the size and item class using a linear mixed effects model with a random intercept for each dog. Fixations, due to our probabilistic approach, could have multiple items. Considering each of these, we had a total of 16,166 individual targets of fixation. Dogs fixated longer on some item classes than others, $\chi^2(14) = 111.49, p < .001$. As an example, dogs' fixations to benches/chairs tended to last for a short time ($M = 166.67$ ms, $SD = 47.14$ ms) but dogs' fixations to cars (another infrequently appearing item) lasted much longer, and were much more variable ($M = 380.87$ ms, $SD = 582.55$ ms). From the same model, we found no main effect of the fixated item size relative to the current view, $\chi 2(1) = .19, p = .66$,

suggesting that dogs' fixations were not significantly longer based on the size of the item they were fixating on. Finally, in the same linear model, we found a significant interaction between item class and fixated item size, $\chi^2(13) = 37.91, p < .001$. This suggests that for some item classes dogs had longer individual fixations in response to larger instances of the item, but that this was not true for all classes. We found little variance explained by our random effect per participant ($ICC = 0.053$).

2.6 Discussion

The present study aimed to explore both the items present in dogs' visual environments as well as which of those items they chose to look at. We integrated computer vision techniques to identify and classify items from the dogs' perspectives. We found that we were able to identify and classify the classes of items available in the dogs' view and how they directed their attention to them. Within our sample, dogs were similar in the item classes they had in their view, mostly encountering plants, buildings, the sky, and pavement. Dogs' visual attention is active and is not just driven by the low level image saliency features (e.g., luminance, color, orientation). Instead, dogs attend selectively to certain item classes.

In our sample, relative to the frequency of the items in their view, dogs fixated proportionally the most to buses, plants, pavement, people, and construction equipment. Of these, buses, people, and construction equipment were all relatively uncommon in dogs' field of view. Dogs looking to plants relatively often was unexpected, as plants are often considered to be background material, similar to the sky which dogs rarely looked to. Further research is needed in an ecologically valid context like the one presented here to further explore what components of plant material dogs are interested in.

We also found suggestions of visual neophilia in dogs, with dogs looking more to unusual and highly variable items, specifically construction equipment. However, dogs were not simply visually attracted to items that are uncommon along our walking route. Infrequently appearing

items along our route that are ubiquitous in dogs' daily lives (i.e., benches and chairs) were rarely looked to. It is unlikely that any of the dogs in our sample were encountering entirely novel item classes, and future research could consider incorporating novel item classes based on dog's past experiences to evaluate if this interest in construction equipment is related to its rarity in dogs' lives or something more tied to the class itself (large brightly colored equipment placed in unusual situations). In our sample, dogs did fixate differently to different classes of items. Some classes like benches and chairs or signs were consistently uninteresting to dogs, but other classes, like people, had significant variation between dogs. The relative consistency in item class presence suggests there may be possible true individual difference in attention to stimuli, including attention to social vs. non-social item classes. This is an interesting interaction, and future work should aim to detect more nuanced individual differences, potentially by recruiting more dogs from a variety of breeds or with more diverse life experiences. Despite this, our findings of individual differences must be treated as preliminary, due to variation in factors such as the time of day, and the presence of mobile item classes (e.g., people, buses) between walks, and variation in the behavior of different dog-guardian dyads (we return to this below).

Like human adults (de Winter et al., 2021; Simpson et al., 2019), dogs frequently attend to classes (i.e., pavement, horizontal plants) that are directly related to ground navigation, helping to find a smooth path and monitor for potential obstacles. Prior work in children (Franchak & Adolph, 2010) has suggested that children also fixate on obstacles in their environment while wayfinding to a similar extent. Future work could explore whether dogs' fixation patterns to obstacles not related to wayfinding in their environment are similar to children, both in their frequency and timing. This would provide interesting comparisons to what is attention grabbing between the two species. Young children and dogs' similar height means that the obstacles and items they encounter in daily life are approached from a similar visual angle, and future work can consider direct comparisons to how they direct their visual attention in complex real-world environments.

Future research can expand on the item classes identified here. Anecdotally, we noticed that

on many fixations to buildings, dogs were specifically looking to the doors and windows of buildings. This pattern was also observed with busses, with dogs often looking at the door of the bus. It's possible that dogs are looking to these portions of buildings and busses because of anticipated functionality (i.e., dogs look to doors of the buildings because they could enter through the door) or for anticipatory social reasons (i.e., people often appear in doorways so doorways are more interesting because of this potential). While this was an interesting observation, due to the limited spatial accuracy of the eye-tracking system given the variable outdoor lighting conditions, we were not able to determine with confidence what component of buildings dogs were looking to on all instances. With increased camera resolution or image enhancement, future work can explicitly identify whether, for instance, dogs look more to the door of buildings and cars than they do to the walls of buildings and bumpers or tires of cars (strictly non-social components). This would provide additional insight into dogs' potential preference for visually attending to social parts of their environment, which could further our understanding of how dogs' complete complex social tasks.

Developing a more nuanced understanding of dogs' attention to social and potentially social areas of their environment would also provide an interesting point of comparison to other non-domesticated canids. While there are no eye-tracking studies in other canid species, prior work comparing dogs' and wolves' gaze behavior has suggested that domestication led to an increased attention to humans and their faces, evidenced in dogs increased looking back behaviors and general visual-social attention (Gácsi et al., 2005; Miklósi et al., 2003). Further, comparative studies can also explore how children and dogs visually engage with social partners on comparable tasks, such as when following a point or working with a partner towards a shared goal. We know that children look to social partners and engage in joint attention tasks (Yu & Smith, 2013) and we could use a similar approach in dogs to explore if their visual engagement patterns are similar.

Reactive dogs (those that have a strong arousal response to a given stimuli such as a dog or a person on a bicycle) would also be promising candidates for use in this paradigm to understand

how they scan their environments and how they, temporally, respond to triggers. Future work can consider applying this method, using what dogs find visually interesting to create a predictive model of behavior for reactive dogs. This could be particularly important for working dogs, predicting suitability for service roles based on how dogs distribute their visual attention. We also found, descriptively, that other features of item classes such as movement were also not large drivers of dog attention. However, many of our item classes moved in a subtle way where the video quality made potential movement challenging to identify. Future work can consider staging instances of identical classes in motion vs. stationary to formally examine the impact of movement on dogs' visual attention.

Another potential source of social input is the dogs' guardian. In our study dog guardians were instructed to walk their dogs as they normally would. No dog guardians actively directed their dogs' visual attention, however there were other differences in the behaviors of dog guardians while walking their dogs. In particular, guardians used different walking styles, varying in how far their dog typically is from them, and in whether they provided their dog with verbal navigational directions. It is possible that these different walking styles may have contributed to differences in dog gaze behavior between different dog-guardian dyads. Limited conclusions can be drawn about these different walking styles at this time given our limited sample. Although we opted to encourage normal walking styles, thereby increasing ecological validity, future work can examine different interaction styles between dogs and their guardians.

Future work could consider exploring how dog guardians looked at their environments in a comparable context. We might expect that both the physical (height) and cognitive differences between the two could result in different patterns of looks. It is also possible that dog owners, when walking their dogs, are following their dogs' gaze and anticipating where they might look. If true, the visual patterns may be quite similar. Future research could also consider running this same study but with human toddlers. Comparable in height to dogs, toddlers may display similar visual patterns to dogs because of their lowered (relative to adults) visual perspective.

While we excluded intensive sniffing bouts where it was unlikely dogs were engaging in

intentional visual processing of the items in their view, there is still the broader question of the relationship between dogs' visual and olfactory attention. Recent work has suggested that dogs' olfaction may be integrated into their visual processing, perhaps into a joint representation of their visual and olfactory environment (Andrews et al., 2022). Therefore, the odors in dogs' environment may be encouraging them to orient their visual attention to a particular visual target or region (and visual attention may similarly be guiding olfaction). In these instances, unlike in our sniffing bouts, attention could still be visual as items are at a distance and reasonable to process visually. Future work can consider examining how dogs integrate olfactory and visual features in a naturalistic context, such as examining how search and rescue or other detection dogs direct their gaze while searching for a person or item.

In this project, we have shown that dogs actively engage with their visual environment. We demonstrated that this approach (head-mounted eye-tracking) can be deployed on a mobile dog to capture ecologically valid data on visual attention. We have also shown that dogs are not looking passively at the environment around them, nor is their looking explained by strictly low-level image properties. With this method, future work can consider exploring what concept spaces dogs may have around item categories. In this project, we identified the items of interest in dogs view using a human concept of "item". Distinguishing between some of these items is likely easy for both species (i.e., there are enormous and meaningful differences between a person and a bench). Selectively looking at certain item classes over others suggests that these differences are detectable by dogs. This study does not reveal what the dog concept or representation of these items is and how their representations differ from our own. Dogs may engage with their visual world in a way that focuses on the affordances of the items present, meaning the ways in which the dog can physically interact with the item. We also know that dogs' olfactory system is highly integrate with their visual one, so they may be focusing on specific items or item categories on the basis of an integrated visual and olfactory cue.

We can now use this method to explore how dogs complete more complex social tasks to better test theories of human-unique cognitive abilities. Little is currently known about how,

visually, dogs build a bond through play with a human companion or how they navigate a complex physical environment with social guidance, such as in dog agility. Understanding the visual behaviors that dogs are utilizing to complete daily tasks, and how those differ from key visual features seen in human-human interactions, will provide insight into social learning and cooperation in a unique cross-species context. Additionally, being able to measure how dogs visually interact with the world while completing tasks is a first step to building an eventual model of dogs' visual behaviors, something with significant implications for both working dog and AI training.

References

Andrews, E. F., Pascalau, R., Horowitz, A., Lawrence, G. M., & Johnson, P. J. (2022). Extensive Connections of the Canine Olfactory Pathway Revealed by Tractography and Dissection. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(33), 6392–6407. https://doi.org/10.1523/JNEUROSCI.2355-21.2022

Bylinskii, Z., Judd, T., Durand, F., Oliva, A., & Torralba, A. (n.d.). MIT Saliency Benchmark. http://saliency.mit.edu/

Bylinskii, Z., Judd, T., Oliva, A., Torralba, A., & Durand, F. (2019). What Do Different Evaluation Metrics Tell Us About Saliency Models? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(3), 740–757. https://doi.org/10.1109/TPAMI.2018.2815601

Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160055. https://doi.org/10.1098/rstb.2016.0055

de Winter, J., Bazilinskyy, P., Wesdorp, D., de Vlam, V., Hopmans, B., Visscher, J., & Dodou, D. (2021). How do pedestrians distribute their visual attention when walking through a parking garage? An eye-tracking study. *Ergonomics*, *64*(6), 793–805. https://doi.org/10.1080/00140139.2020.1862310

Franchak, J. M., & Adolph, K. E. (2010). Visually guided navigation: Head-mounted eye-tracking of natural locomotion in children and adults. *Vision Research*, *50*(24), 2766–2774. https://doi.org/10.1016/j.visres.2010.09.024

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011). Head-Mounted Eye Tracking: A New Method to Describe Infant Looking: Head-Mounted Eye Tracking. *Child Development*, *82*(6), 1738–1750. https://doi.org/10.1111/j.1467-8624.2011.01670.x

Gácsi, M., Győri, B., Miklósi, Á., Virányi, Z., Kubinyi, E., Topál, J., & Csányi, V. (2005). Species-specific differences and similarities in the behavior of hand-raised dog and wolf pups in social situations with humans. *Developmental Psychobiology*, *47*(2), 111–122. https://doi.org/10.1002/dev.20082

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018). Mask R-CNN. https://doi.org/10.48550/arXiv.1703.06870

Hunnius, S., & Bekkering, H. (2010). The early development of object knowledge: A study of infants' visual anticipations during action observation. *Developmental Psychology*, *46*(2), 446–454. https://doi.org/10.1037/a0016543

Jayaraman, S., & Smith, L. B. (2020). The Infant's Visual World: The Everyday Statistics for Visual Learning. In J. J. Lockman & C. S. Tamis-LeMonda (Eds.), *The Cambridge Handbook of Infant Development* (1st ed., pp. 549–576). Cambridge University Press. https://doi.org/10.1017/9781108351959.020

Judd, T., Durand, F., & Torralba, A. (2012). A Benchmark of Computational Models of Saliency to Predict Human Fixations. *MIT Technical Report*.

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and Walking Infants See the World Differently. *Child Development*, *85*(4), 1503–1518. https://doi.org/10.1111/cdev.12206

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2015). *Microsoft COCO: Common Objects in Context* (tech. rep. No. arXiv:1405.0312). arXiv. http://arxiv.org/abs/1405.0312

Miklósi, Á., Kubinyi, E., Topál, J., Gácsi, M., Virányi, Z., & Csányi, V. (2003). A Simple Reason for a Big Difference: Wolves Do Not Look Back at Humans, but Dogs Do. *Current Biology*, *13*(9), 763–766. https://doi.org/10.1016/S0960-9822(03)00263-X

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022). Head-mounted mobile eye-tracking in the domestic dog: A new method. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01907-3

Reynolds, G. D. (2015). Infant visual attention and object recognition. *Behavioural Brain Research*, *285*, 34–43. https://doi.org/10.1016/j.bbr.2015.01.015

Rodin, I., Furnari, A., Mavroeidis, D., & Farinella, G. M. (2021). Predicting the future from first person (egocentric) vision: A survey. *Computer Vision and Image Understanding*, *211*, 103252. https://doi.org/10.1016/j.cviu.2021.103252

Simpson, J., Thwaites, K., & Freeth, M. (2019). Understanding Visual Engagement with Urban Street Edges along Non-Pedestrianised and Pedestrianised Streets Using Mobile Eye-Tracking. *Sustainability*, *11*(15), 4251. https://doi.org/10.3390/su11154251

Smith, L. B., Jayaraman, S., Clerkin, E., & Yu, C. (2018). The developing infant creates a curriculum for statistical learning. *Trends in cognitive sciences*, *22*(4), 325–336. https://doi.org/10.1016/j.tics.2018.02.004

Smith, L. B., Yu, C., Yoshida, H., & Fausey, C. M. (2015). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. *Journal of Cognition and Development*, *16*(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

Törnqvist, H., Somppi, S., Kujala, M. V., & Vainio, O. (2020). Observing animals and humans: Dogs target their gaze to the biological information in natural scenes. *PeerJ*, *8*, e10341. https://doi.org/10.7717/peerj.10341

Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

Watalingam, R. D., Richetelli, N., Pelz, J. B., & Speir, J. A. (2017). Eye tracking to evaluate evidence recognition in crime scene investigations. *Forensic Science International*, *280*, 64–80. https://doi.org/10.1016/j.forsciint.2017.08.012

Yorzinski, J. L., Patricelli, G. L., Babcock, J. S., Pearson, J. M., & Platt, M. L. (2013). Through their eyes: Selective attention in peahens during courtship. *Journal of Experimental Biology*, *216*(16), 3035–3046. https://doi.org/10.1242/jeb.087338

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

CHAPTER 3

Point Comprehension Across Species [1]

3.1 Background

How are pointing gestures interpreted? Are the cognitive mechanisms that allow toddlers to follow points unique to our species? Pointing is one of the most widely studied gestures. Depending on your theoretical perspective, pointing may be a rich and collaborative communication of shared intent, or it may be an association with a specific body position learned through extensive exposure. An existing challenge is distinguishing between what is necessary to solve the task at hand, and what is actually happening in the mind. Regardless of the interpretation, point following requires the viewer of the point to notice that a gesture is being made, and to respond to that gesture in a context-appropriate fashion.

3.1.1 Lean Interpretations

It is possible that point following in young children can be explained by low-level associative mechanisms. Theories of associative learning have provided a robust source of research for over 125 years. These theories have changed since the days of Pavlov, but most associative

[1] This chapter contains a significant expansion and extension of the analysis approach used in Pelgrim et al. (2024) published in the Proceedings of the Cognitive Science Society

theories follow a similar structure. In general, associative theories view learning as the pairing or connection between previously unrelated things, typically a stimulus of some kind and a response. This learned connection may result in increases or decreases in behaviors or responses, depending on the specific paradigm. The response may be voluntary, as is the case in operant conditioning, where the learner pairs their actions with some outcome. Operant conditioning can explain learned behaviors ranging from simple, like pressing a lever, all the way up to extended chains of complex learned actions. Using this mechanism children may learn to follow a point, or follow the gaze of a pointer (Corkum & Moore, 1998). A child may initially look at the target indicated by the pointer by accident or by chance. Once they "follow" the point, they are rewarded with an enriching social interaction. They may even get to play with the target of the point. This social feedback and play acts as a positive reinforcer, increasing the probability that the child will perform the previously rewarded behavior in the future. Although the underlying mechanism of pairing is simple, associative mechanisms can be used to explain complex behaviors and learning patterns, like avoidance learning or imitation (Lind et al., 2019).

In some sense, associative learning (like the ability for toddlers to follow a point) is a settled idea. There is little doubt that learning can and does occur via the proposed associative mechanisms. In domains beyond point following, children selectively learn word labels for novel objects from individuals who were previously accurate at labeling familiar objects, over those who were previously inaccurate at labeling familiar objects (Birch et al., 2008; Koenig & Harris, 2005; Pasquini et al., 2007). This is a well-established finding and one that is plausible under associative mechanisms. Children could learn that one informant tends to be correct (relative to the other) after being exposed to accurate and inaccurate informants labeling familiar objects. Further, because associative mechanisms are domain general, children would be expected to generalize this learned knowledge of accuracy across other domains. This generalization also termed a halo effect, is found in children. Children expect previously accurate informants to be more knowledgeable and prosocial (Brosseau-Liard & Birch, 2010).

Lean associative accounts can also explain many of the ways that dogs interact with human

beings. Puppies who live in homes with humans outperform puppies who live in shelters by following more difficult points (Zaine et al., 2015). Zaine et al. (2015) argue that this improvement is due to increased exposure to humans, facilitating associative learning between human pointing and a reward.

Associative learning mechanisms are also plausible explanations for how dogs are able to evaluate individuals beyond point comprehension. Dogs are able to make judgments about individuals based on observing the individual's interaction with a third party. Also known as social eavesdropping tasks, in these studies, dogs observe two human informants interact with a third person. One informant (the generous individual) gives food to the third person, while the other informant (the selfish individual) does not. Both informants then hold out pieces of food, and the dog is allowed to choose which person they want to eat the food from. In these studies, dogs choose the previously cooperative or generous person (Kundey et al., 2011; Marshall-Pescini et al., 2011). Critically, when the informants change locations after the observed interaction with the third party but before the dog's choice (such that the generous individual is no longer always on the same side as the food transfer), dogs no longer prefer the previously generous individual (Nitzschner et al., 2014). This provides support for associative or lean perspectives, as dogs were going to the location that was previously associated with food rather than making inferences about the individuals.

However, there is still significant debate over the extent to which associative learning underlies point following and social learning more generally. Children are sensitive to the area of expertise of an informant, and confine that informant's trustworthiness to the domain they are an expert in (Koenig & Jaswal, 2011; Sobel & Corriveau, 2010; Sobel & Kushnir, 2013). They will preferentially learn from (or ask to learn from) the expert informant when the problem they need help with is in that expert's domain. In contrast, for questions in an unrelated subject area, children do not show the same expert preference. As an example, when presented with two informants, one of whom is a dog expert, children prefer learning from the dog expert when presented with novel dogs whose breed they do not know. Critically, they choose equally

between the informants when presented with novel objects that neither informant is a known expert in (Koenig & Jaswal, 2011).

In explicit tests of selective trust towards human informants, dogs' behaviors are difficult to explain via associative mechanisms alone. As an example, dogs are sensitive to the knowledge of an informant in guesser-knower paradigms. In these set-ups, one informant (the knower) either hides or watches while a third person hides a treat. A second informant (the guesser) does not observe or participate in the hiding. The dog is not able to see where the treat is hidden. The informants then both uncover their faces and point at conflicting locations in an identical manner. Dogs selectively follow the point of the knowledgeable informant, choosing the location indicated by the knower more often than the location pointed to by the guesser (Catala et al., 2017; Maginnity & Grace, 2014). Dogs continue to perform successfully at guesser-knower paradigms when associative cues like food handling, (controlled for by having a third person bait so the knower does not handle the food), or trial-and-error learning (controlled for by examining first-trial performance and checking for learning across trials) are minimized.

Lean interpretations of point following are challenged by tests of point following on the basis of prior informant accuracy. An ability dogs appear to share with human children, both species are able to use past informant accuracy to selectively trust, choosing the location indicated by a previously accurate individual significantly more than the location pointed to by a previously inaccurate person (Pelgrim et al., 2021). Unlike human children, it is not currently known how dogs will respond to probabilistic informants. More specifically, dogs are able to selectively trust informants on a binary task (trusting an informant who was 100% correct vs. an informant who was 100% wrong). However, no study in dogs has examined informants with varying degrees of accuracy (i.e., 75% correct vs. 25% correct or 75% vs. 0%) as has been done in children (Pasquini et al., 2007). 4-year-old children can use probabilistic accuracy to selectively trust an informant who was previously wrong (e.g., 75% correct), but who was more accurate than the other informant (e.g., 25% correct). 3-year-olds are unable to selectively trust informants based on past accuracy when that accuracy is probabilistic, such that both informants were wrong at

least once. 3-year-old only succeeded when one informant was 100% correct and the other was 25% correct (Pasquini et al., 2007). If associative learning was the sole mechanism for pointing comprehension, children should not have a complete avoidance of informants who make errors at age 3, and then lose this bias a year later.

3.1.2 Rich Interpretations

It is possible that pointing in young children and in other species is a rich social-communicative gesture. Points could serve as bids for joint engagement, and as a way to collaboratively engage in shared goals and intentions without necessitating language. If true, the point is a uniquely human gesture that facilitates joint attention (Butterworth, 2003; Tomasello et al., 2007).

Under the view of rich pointing interpretations, pointing without rich "what" and "why" mental explanations is a fundamentally underdetermined problem (Tomasello et al., 2007). In real-world contexts, points are frequently used to refer to specific sections of larger overlapping objects. In the context of word learning, even into adulthood, identifying what is being referred to in a point is critical. For example, a learner is playing chess for the first time. Their teacher points with the accompanied verbal cue "this is the bishop". Without an understanding of "why" and "what" this gesture is ambiguous. To make sense of this verbal label and gesture, the learner must know that the bishop is a piece in the game they are preparing to play, that the small figurine is representing the piece we refer to as the bishop, and that the term bishop does not refer to any of the adjacent pieces or the board itself (Wittgenstein, 1989). All of this must be known by our learner, and our learner must also know it is known by the pointer. The pointer and learner, in this chess example, are both aware of the shared goal of the situation (teach/learn how to play chess), the context and background of the participants (chess expert/chess novice) and the boundaries around the kinds of information being provided (provide what must be known/attend to information being provided to acquire the knowledge to play the game).

Understanding the "what" and the "why" helps to explain infants success at relatively ambiguous pointing gestures. When pointing gestures are produced accidentally or in a distracted

fashion, children as young as 14 months do not preferentially follow them. In contrast, in the same paradigm when pointing cues are produced with communicative intent, including facial expressions and gaze alternations, the same age group followed the point to locate the hidden item (Behne et al., 2005). Children are able to interpret points as being directed for them by detecting ostensive cues. Ostensive cues are social gestures, expressions, and engagement styles that help the viewer know that the message is intentional and for them. These cues can include facial expressions, such as raising the eyebrows or smiling, as well as eye contact. Ostensive cues can also extend beyond the visual domain, using vocalizations in different pitches or tones, as well as calling the participant's name. Recent findings also suggest that dogs may not be as sensitive to ostensive cues as human children. In a large multi-lab replication site, dogs did not follow pointing gestures better when they were delivered ostensively than when they were delivered non-ostensively (ManyDogs et al., 2023). This finding does have some significant limitations acknowledged by the authors themselves (high participant dropout rate, protocol variations between labs, data collection during COVID-19 requiring face masks, etc.). However, it suggests that there may be richer cognitive abilities necessary for human-like point following.

To understand pointing in a rich way, there are a variety of cognitive abilities that must be present. Rich interpretations of point following often follow a sort of dual-process approach, arguing that point following prior to 12-14 months of age is explained by a lower-level or associative mechanism. Once all the necessary cognitive prerequisites are present in the infant, they interpret pointing as a highly social gesture with layers of implicit information. These prerequisites include the understanding of goals of others, knowledge of others, and intentions of others (Tomasello et al., 2007).

Rich interpretations of point following also help to explain discontinuities observed in child development. Very young children display a default bias to trust, and this can be explained as operating on the low-level innate or early-learned mechanism (Jaswal et al., 2010). As children develop, they acquire cognitive resources (e.g., improved executive function, more knowledge of word meanings and linguistic rules), a more critical stance, and are better able to reason

about the speaker's intentions, resulting in changes in their trust in others (Harris et al., 2018; Mills, 2013; Sobel & Finiasz, 2020; Sobel & Kushnir, 2013). This development also helps to explain why young children find it challenging to select optimally between informants who are probabilistically accurate (Pasquini et al., 2007).

In dog developmental paradigms, puppies are able to follow points from the earliest age that they can be tested (Agnetta et al., 2000; Hare et al., 2002)[2]. Given that these puppies have no explicit training and have had very limited exposure to humans (in part due to their sensory limitations), this suggests some level of innate sensitivity to human social communicative gestures that are difficult to explain through associative learning. Dog pups also outperform wolf pups at following subtle human pointing gestures, even when the wolf pups have been hand raised and intensively socialized with humans (Virányi et al., 2008).

Children are capable of using "higher-order" mechanisms to make decisions and solve problems, but is this their default mode when interpreting point following? Classic studies have made it challenging to detect.

3.1.3 Testing Pointing Comprehension

Pointing comprehension and production have often been tested as a function of the style of point produced, as discussed in Chapter 1. Pointing can be a declarative, directing the viewer's attention to an item. Pointing can also be a social tool, acting as an imperative to make a request or command (Bates et al., 1975). Imperative pointing can also serve a cooperative goal, requesting assistance when working together (Tomasello et al., 2007). We can also point to request information from others. Termed interrogative pointing, this style serves as an epistemic request (Begus & Southgate, 2012; Kovács et al., 2014; Southgate et al., 2007). Across cultures, 12-month-old infants can both follow and produce these kinds of pointing gestures (Liszkowski

[2]Unlike humans, dogs cannot participate in cognitive studies as newborns to control for any neonatal learning (at least not as behavioral studies are currently conducted). Puppies are born deaf, blind, and with very limited motor abilities (they also likely are not able to fully smell). Dogs are not able to fully hear (ears open and functional) and see (eyes completely open with pupillary reaction) until around 3-4 weeks of age (Scott, 1958)

et al., 2012).

Testing point following, or point comprehension, classically involved participants being presented with a point and multiple targets. Younger infants are often tested using looking time based paradigms. As children grow and are more mobile, their ability to comprehend points is frequently tested in search paradigms.

Classic tests of point comprehension with mobile participants require the participant to search for a target hidden in one of two locations. The person who produces the point (hereafter referred to as the pointer) generally stands between the two search locations and executes a pointing gesture. As mentioned previously in this dissertation, children are easily able to understand and follow points from a young age, selecting the indicated location more than would be expected from random chance (Aureli et al., 2009; Pfandler et al., 2013). However, classic studies of pointing have been limited by an inability to differentiate interpretations of how the pointing gesture is processed, and whether this differs by species.

We can walk through the implications of multiple predictions in the case of a classic experimental paradigm, a toddler searching for a hidden object while watching another person deliver a point. It is possible that young children understand that they are receiving information that will help to facilitate a joint goal or intention as part of a broader game/context. If young children generally perceive points as the start of a collaborative task involving shared engagement, our participant should be able to infer additional information from the situation. They understand that the pointer is directing them to a specific referent (the target) in the environment. They also understand that, in the context of a search task, the broader goal of the game is for them to approach the target, and only the target. This is all facilitated by cooperating with the pointer and attending to the cues they are intentionally producing. We might expect to see instances of joint attention and mutual gaze, and that our toddler should be resoundingly successful at locating the target of the point. It is also possible that alternative cognitive processes could lead to the same observed behavioral results. Maybe toddlers understand points as a broad directional cue, indicating that something of interest is in the region indicated by the pointer. In the context of a

classic two-alternative forced choice task, these two cognitive mechanisms are indistinguishable. Our participant could successfully "get the point" and select the correct location by following associative cues. This suggests a need for a more complex task.

Dogs likewise present an interesting opportunity to gain a more nuanced understanding of how pointing gestures are processed. Recent work with more complex pointing gestures in dogs, including the elimination of gazing cues by having the pointer stare directly at the ground or the dog, has suggested that dog point following may be more brittle than previously thought (Eatherington et al., 2021; Lakatos et al., 2012; ManyDogs et al., 2023). In particular, when the pointer looks in a different direction than the location they are pointing, and when dogs do not have access to details about the direction of the head and gaze of the pointer (through the use of a silhouette pointing rather than a 2D video of a person), dogs perform much more poorly. Dogs' recent poor performance on a subset of point following tasks has created disagreement in the literature and left our understanding of dogs' response to human social-communicative gestures unclear.

Dogs and young children provide a particularly helpful comparison as, unlike other species like non-human primates, pet dogs live in and are raised in our homes. Comparing dogs and children on comparable behavioral tasks has helped to better understand a variety of topics, including overimitation, point following, and approaches to unsolvable tasks (Fisher et al., 2025; Lakatos et al., 2009; Marshall-Pescini et al., 2013). This provides further opportunities to understand the role of point following in a more complex task as it relates to human-unique cognition (Lyn & Christopher, 2020).

In addition to comparing species and using a more complex task, one way to better understand the components of human pointing is taking a more quantitative approach. We can describe points based on their function (Bates et al., 1975), based on their relative distance (Lyn & Christopher, 2020) or based on the body gesture itself. In the developmental literature, this has often focused on whether children display the canonical index finger point, or whether they point with their whole arm. However, particularly when pointing to relatively distant referents,

pointers move their bodies and not just their hands.

Independent of the cognitive mechanism underpinning point following, the gesture itself contains a rich suite of potential cues that could help the point follower identify the referent. Adult humans are easily able to produce and follow pointing gestures in a sophisticated fashion. When following pointing gestures in crowded or complex environments, we may not always have the rich kind of "what" and "why" information that can help identify the referent. Pointing following in adults is often tested by requiring the observer to indicate a specific square or item in a complex grid that the pointer is indicating. These items are typically adjacent, leaving little room for error. To succeed, adults have to account for the perspective of the observer of their point. Adults are able to solve these highly complex tasks, but they do sometimes make mistakes. Pointers can systematically undercorrect or overcorrect for the perspective of the observer, leading them to an adjacent item. Point followers can make the same perspective taking failures, misjudging the target of the pointer (Herbort & Kunde, 2016; Herbort et al., 2021). Given the challenges that adults face in solving the geometry of the problem of identifying pointing referents, it is worthwhile to consider quantitatively the way that pointers dynamically move their bodies.

Taking a quantitative approach helps us to evaluate these body language cues, revealing how potentially ambiguous or clear following specific body parts may be. To do this we can take inspiration from computer science and robotics. Unlike in past work with children and dogs, where points are typically described qualitatively, point following in computer science and robotics is programmed using computer vision and mathematical analysis of candidate cues. This makes the evaluation of varying sources of information from a point gesture quantitative (Azari et al., 2019; Hu et al., 2022; Trenkle et al., 2015).

To interpret a pointing gesture quantitatively, you must first identify where the pointer is in the room (typically using computer vision). Then, you must identify key body parts on the person, such as the elbow and the eyes. Finally, you can project into the environment a line extending into space from the candidate body language cues. In robotics, this has been

termed ray casting vectors. Multiple candidate vectors have been explored for human pointing, particularly following the line from the elbow to the hand, from the shoulder to hand, or from the head to the hand (Abidi et al., 2013; Hu et al., 2022; Taylor & McCloskey, 1988). By quantifying how experimenters produce points, we can examine what the best-performing vectors are and if they differ by location. Further, we can compare how dog owners naturally point for their dogs (Chapter 4) to better capture if the kinds of points we use in studies are comparable to those dogs see in their daily lives.

We can take further inspiration from perceptual studies to generate a more nuanced understanding of how our participants are assessing the point and making their choices. Tracking trajectories of movement has led to significant insights across a variety of modalities. In seated tasks, or those requiring arm reaching, we know how competing cues are processed dynamically over time (Erb et al., 2016; Song & Nakayama, 2009). Tracking trajectory in walking tasks can also reveal changes in visual perception as we walk in real and in virtual environments, navigating and avoiding obstacles (Cirio et al., 2013; Fink et al., 2007).

In Experiment 1, we presented dogs and toddlers with a search task. We required them to follow a pointer to locate a target hidden inside one of four items. Our task added a dimension of difficulty by expanding the potential search items from two (as in classic studies) to four, spread equally in front of the participant. By including multiple items (two) on each side of the body, we could better understand how the pointing gesture itself was processed and if there were species differences. If pointing is interpreted as a broad body-side cue, we would see no differentiation between the inner and outer locations on either side of the experimenter. Alternatively, if pointing is interpreted by toddlers as a collaborative engagement task, they should localize the referent and perform similarly to classic two item studies.

We recorded depth data throughout Experiment 1 to quantify the pointing gestures produced for dogs and toddlers by our experimenters. This provided a comparison point for how dog owners point (see Chapter 4) and helped us to better understand what kinds of body cues may be best to follow. We also measured the path taken by participants when making choices. This

helped to provide a more nuanced understanding of their point comprehension beyond their choice behaviors. If participants were confident of the location of the item, we would expect them to walk directly to that item, taking close to the most efficient route to get there. In contrast, if they are less confident or are weighing competing cues in the location of the item, they may take a more variable route to their first choices. We examined if this differed by species.

In Experiment 2, we presented adults with an online version of the same task. This allowed us to confirm that the task was easily solvable by an adult population. Both Experiments were pre-registered.

3.2 Experiment 1: Toddlers & Dogs

3.2.1 Participants

The participants were 26 dogs (13 = Female, 13 = Male, Mean Age = 59.36 Months, Age Range = 9.85 - 148.49) and 26 children (11 = Female, 15 = Male, Mean Age = 19.67 Months, Age Range = 17.23 - 24.85). An additional 2 dogs were excluded due to (1) Experimenter error (N = 1), and (2) repeated no-choices (N = 1). An additional 5 children were excluded due to (1) Experimenter error (N = 1), and (2) unwillingness to complete all trials (N = 4).

3.2.2 Ethics Note

All ethics guidelines were followed, and study procedures were approved by the relevant ethics committees. All components of the study involving dogs were approved by the Brown Institutional Animal Care and Use Committee, Protocol Number 23-09-0002. Components of the study involving children were approved by the Brown Institutional Review Board Protocol Number 2105002985.

3.2.3 Procedure

During our experiment, dogs searched for a hidden treat, and children searched for a hidden ball.

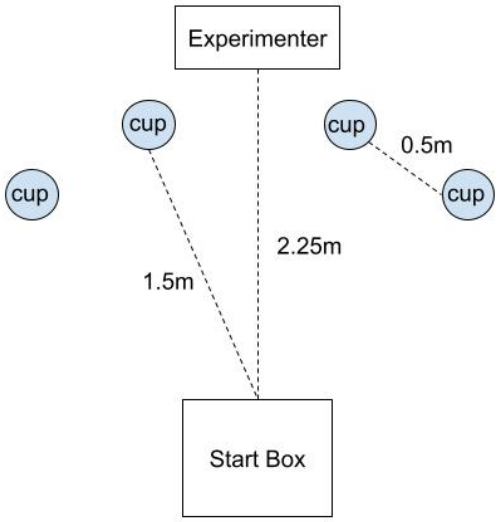

Figure 3.1: Room set-up used during the study.

Warm-Up Activities

Prior to starting the experiment, dogs engaged in a warm-up activity to receive a treat at each of the 4 hidden search locations with baiting visible. Dogs watched the experimenter place an empty cup onto the ground to serve as a distractor. The experimenter then said "[dog name], look" while showing the dog the treat and holding a second cup. They then placed the treat into the second cup and placed the cup onto the opposite location to where the first cup already was (i.e., if the first cup was at the outer right location the second cup would be placed on the outer left). Dogs were then released to search for the treat, and were assisted in getting the treat out of the lidded cup by the experimenter if required.

Before test trials, toddlers were familiarized with the jingle box, a custom-built box that makes music when a ball is dropped into it. To improve engagement for toddlers who were not

motivated to search for the ball or engage with the jingle box, toddlers were also introduced to a stuffed dog called Bruno. Toddlers were instructed that they would help Bruno find his ball, and that the experimenter would hide it. Toddlers were generally familiar with the cups being used from their daily lives (silicone snack cups with integrated lids). Still, they were shown in the same manner as the dogs that the ball could be hidden and retrieved from the cup by the experimenter prior to starting our trials.

Test Trials

For each trial, the participant waited with their guardian (parent or dog owner) approximately 2.25 m away from the experimenter in the start box (Figure 3.1). Dogs waited in the box while their owner handled their leash in a chair immediately behind them. Toddlers waited in the box with their parent, while either sitting on the parent's lap or sitting next to the parent.

A trial began when the experimenter said "[participant name], look" while showing the search item. The experimenter then turned their back and hid the item inside one of four silicone-lidded cups on a small table behind them. The experimenter then placed the cups on the search locations (order semi-randomized). Cups were numbered 1-4, starting with the experimenter's outer left side (cup 1). Cup 2 was on the inner left, cup 3 was on the experimenter's inner right, and cup 4 was on the far right.

After placement, the experimenter looked at the participant with both arms by their sides and said "[participant name], look". The experimenter then pointed to the location of the hidden item using the proximal arm. After 2 seconds, the experimenter said "OK" which cued the guardian to release the participant to approach.

Participants could search exhaustively. Choice was considered physical contact with the cup, and the order of choices was recorded. The experimenter maintained their point at the hidden item for the duration of the trial, except if the participant had difficulty opening or searching the cup (i.e., dog with a flat face did not have a muzzle to push inside the cup) or removing the cup from their body (i.e., cup stuck on a child's arm or dog's nose). In these cases, the

experimenter stopped pointing and assisted the participant to remove the empty cup from their body or revealed that the cup was empty. The experimenter then resumed pointing and verbally encouraged the participant to locate the hidden object (i.e.,"Where is the [treat/ball]".

Once the participant found the hidden object, the experimenter stopped pointing and assisted the participant in removing the object if required. Dogs were allowed to eat the treat and were then recalled by their guardians. Toddlers returned the ball to the jingle box or otherwise played with the ball. After each trial, participants were praised by the experimenter and/or guardian regardless of performance. This procedure was repeated for a total of 10 trials. The location of the hidden object was semi-randomized so that participants saw a total of 5 trials to the two locations closest to the experimenter and 5 to the locations farthest from the experimenter. Objects were also hidden an equal number of times on the left and right sides of the experimenter.

Across the test trials, we used two Intel RealSense cameras to record both depth and RGB data. This allowed us to track the experimenter's body positions and quantify their pointing, helping to better understand what kind of gestural cues were available to participants (Pelgrim et al., 2024). We also tracked the participants' movements during their choices, capturing their position as they approached.

3.2.4 Data Coding and Preparation

Behavioral Data

As mentioned above, choice was defined as physical contact with the cup. Participants were permitted to search exhaustively, and the order of choices was recorded live by a researcher during the session. We also noted the number of searches required to locate the hidden object. In the event that the participant searched but did not end up locating the hidden object (1/260 trials for dogs and 3/260 trials for toddlers), we excluded that trial from analyses on the number of trials required to locate the object.

A subset of participants (25% for each species or $7/26$) were re-coded from video recordings

of the session by a research assistant who had not participated in the original sessions. Inter-rater reliability was exceptionally high for both dogs and toddlers. For each species, coders agreed on 69/70 (98.6%) duplicate-coded trials. In the four instances of disagreement (two for each species), I reviewed the re-codings and decided upon the final version.

Motion Tracking

To process our data we had to identify three elements: the search targets (the four cups at fixed locations), the pointer (the experimenter standing at a fixed location), and the dog or toddler (moving from the start marker to search during the trial). We began by segmenting the footage to identify each of the three listed elements (cups, experimenter, participant) using YOLOv8 (Jocher et al., 2023). After identifying where the experimenter was, we estimated the key locations on the body. We used Google's MediaPipe Pose Landmarker (Bazarevsky et al., 2020) to process videos of the pointing gestures. We then used the depth information to transform the coordinates of the relevant key body parts from 2D to 3D space (Figure 3.2).

Using the 3D representation of the person, we ray-casted our candidate vectors from the experimenter's body. Experimenters always pointed with their lateral or same-side arm. This made identification of the pointing arm easy via manual specification based on target location.

In keeping with prior work, we initially evaluated five pointing vectors. First, we used the Eye-Wrist, defined by a vector connecting the eye and wrist of the pointing arm. Next, we use the Nose-Wrist, which is defined by a vector connecting the nose and wrist of the pointing arm. Third, we use the Shoulder-Wrist, a vector representing the arm, connecting the shoulder and wrist. Fourth, we use the Forearm or Elbow-Wrist, defined by a vector connecting the elbow and wrist of the pointing arm. Finally, we explored using the Gaze, a vector representing the general direction the user is looking at. To find the gaze vector, we computed the normal vector to the plane passing through their left eye, right eye, and center of the mouth. In our sample, we had difficulty tracking the experimenters' faces throughout the whole trial with sufficient accuracy to generate the gaze vector. Given this inaccuracy, and the fact that the gaze vector

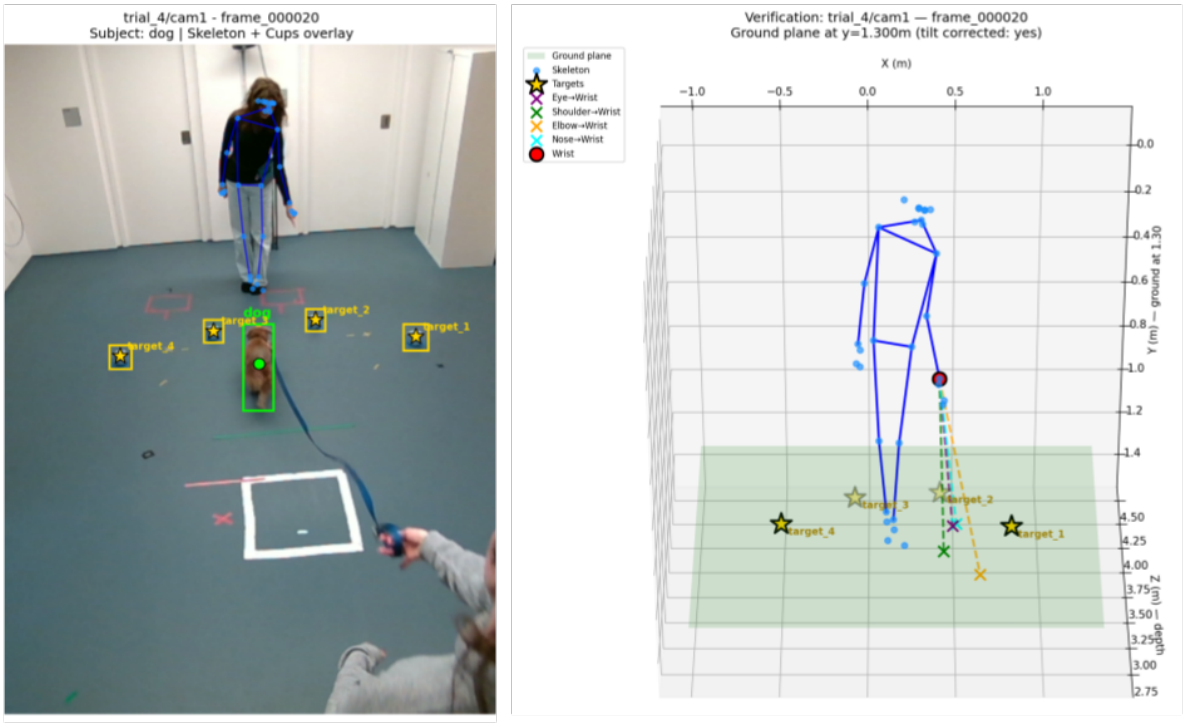

Figure 3.2: **Left**: post-processing output with the cups, dog, and person's key body points identified. **Right**: pointing vectors from the image on the left projected into space.

has been demonstrated to perform poorly as a localizer of a pointing target in naturalistic point production, we opted to proceed without the gaze vector (Pelgrim et al., 2024).

We used three metrics to quantify the candidate vectors generated from the points. First, for each of our 4 vectors, we calculated the Euclidean distance from where the vector intersected the ground to the target cup. Second, we calculated the weighted accuracy. Weighted Accuracy, where A is 1 if the correct target was selected and 0 otherwise, n is the number of selections made until the target is selected, and w is the probability of a target being selected, was calculated using the normalized inverse Euclidean distance.

$$acc = \frac{\sum_{i=1}^{n} w_i A_i}{\sum_{i=1}^{n} w_i} \tag{3.1}$$

Finally, we evaluated uncertainty using a measure of perplexity. Perplexity metrics have previously been widely used in language or speech recognition tasks (Jelinek et al., 2005; Wang et al., 2023). Perplexity measures the degree of uncertainty in the value of a sample from a discrete probability distribution. The larger the perplexity, the less likely it is that an observer can guess the value, or in our context, the target of the point, which is drawn from the distribution. To calculate gestural perplexity, our work assumes the pointing vectors can be parameterized by two angles, corresponding to the "pitch" and "yaw" of the pointing vectors, which we term α and β. We assume the vectors are parameterized by the person's location, which is known, height, which is directly observed, and two angles, α and β, giving us:

$$\Pr(o|s,a) = \Pr(\alpha,\beta|s,a) \tag{3.2}$$

For each instance of pointing, n, we have information about the true location of the target of the point, X, as well as the pointing ray α and β, observed from depth measurements. In the case of our trials, where we know the object is in one of k predefined locations t_i and the distance from the pointing ray intersection location to each target d_j, we can compute the perplexity over the dataset N as a multinomial over the true location as follows:

$$Perplexity(N) = exp\left\{-\frac{1}{|N|}\sum_{n\in N}\log \mathscr{L}(t_i|d_1,...,d_n)\right\} \tag{3.3}$$

Participant Path Data

To track participants' movement in space, we used an off-the-shelf computer vision algorithm (SAM3) that automatically detected the location of our participants for each trial (Carion et al., 2025). Once the location was integrated with our depth data, we determined the proximity of the participant to each of our possible search locations as well as to the experimenter. This allowed us to examine participant location and generate their overall path towards the targets. For participants' first choice, we analyzed the average deviation from the most direct path by

species and by location. This captures how variable the paths were to participants' first choices. We also reported on the maximum deviation from the most direct path to the location of the cup they first chose.

3.2.5 Results and Discussion

Behavioral Data

In accordance with our pre-registration, we first examined if dogs and children as groups were locating the hidden object on the first trial at above chance (25%) levels using a one-sample t-test. Dogs located the treat on the first try at above chance (0.25) levels, M = 0.397, 95% CI [0.32, 0.47], $t(25) = 4.09, p < .001$. Of our 26 dogs, 20 (76.9%) located the treat on the first try at above chance levels. Similarly, toddlers located the hidden toy on their first choice at above chance (0.25) levels, M = 0.532, 95% CI [0.45, 0.61], $t(25) = 7.33, p < .001$. All but one of the toddlers in our sample (96.2%) located the toy on their first choice at above chance (25%) levels. It is interesting that although participants are numerically above chance levels, they are nowhere near ceiling level, Figure 3.3.

We also examined whether dogs and toddlers were performing differently from each other, examining the proportion of trials in which they located the object on the first try using an independent samples t-test. Toddlers found the hidden item on the first try more often than dogs, $t(49.79) = 2.56, p = .014$. Both species frequently approached empty cups, suggesting that solving the problem of both identifying the object referent and inhibiting approach to all other objects is a challenging task.

In line with our pre-registration, we also descriptively examined participants' accuracy based on whether they first chose one of the two cups on the correct side of the experimenter (the Left or Right), ignoring whether among the two cups they chose the true baited one. To do this, we collapsed choices of cup 1 and 2 (the two cups on the left side of the experimenter) and choices of cup 3 and 4 (the two cups on the right side of the experimenter). When averaging by the side

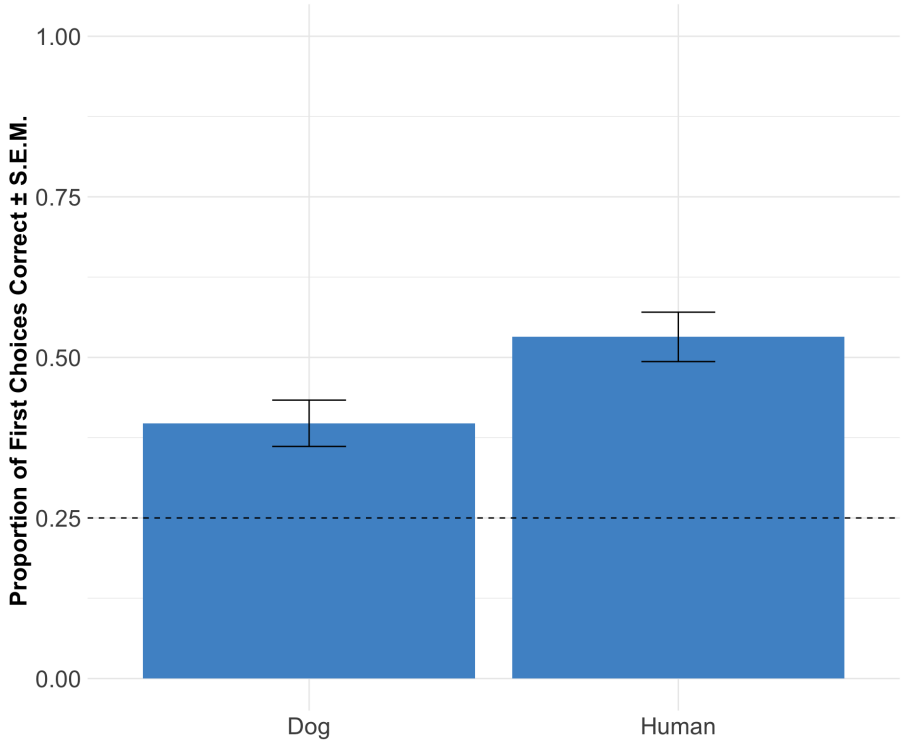

Figure 3.3: Proportion of first choices correct by species.

of the experimenter, kids and dogs both performed above chance (50%) levels. Dogs chose the correct side of the room 73% of the time, and toddlers chose the correct side 75% of the time. Interestingly, dogs and toddlers differed in how much they preferred the inner and outer cups. Dogs appeared to be attracted to the experimenter, choosing the inner locations 73% of the time. In contrast, toddlers chose between the four locations approximately equally. See Figure 3.4. Neither species displayed a side bias on average. Both dogs and toddlers chose approximately equally to the left and right sides of the experimenter.

The pattern of attraction to the inner cups helps to explain how toddlers outperform dogs when we examine the choice of the exact correct item, but why that advantage disappears when we average across room sides. More specifically, dogs are attracted to the inner cups (cups 2 and 3) closest to the experimenter and in turn make a systematic error of biased interior choices. In contrast, kids display no such preference and make uniform choices. Toddlers also do not display any side biases. This pattern of results may help to explain the different point following or point comprehension abilities between the two species. Dogs appear to factor in the proximity

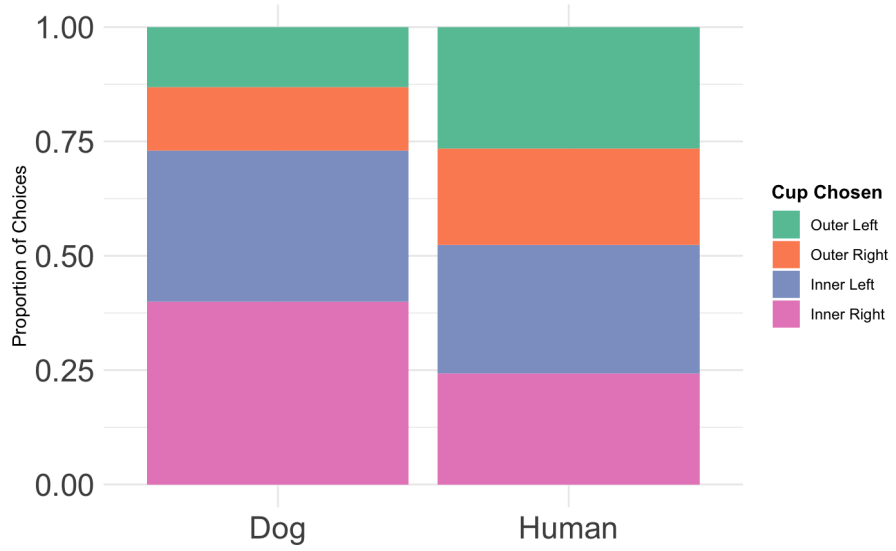

Figure 3.4: Proportion of first choices to each of the 4 cups by species.

of the experimenter as well as the pointing gesture they are using. For example, dogs may attempt to consider the pointing gesture itself as one factor, but they also have a second factor, the experimenter's location, that combines to down-weight the choice value of the two outer locations. It is also possible that dogs are following a low-level choice rule, such as go to the side with the arm extended selecting the closest cup to the experimenter. Toddlers appear to only account for pointing gestures, and from their choice data it is not obvious why errors are happening. To gain a more nuanced understanding of how the choices were being made, rather than just their choices themselves, we can examine the participant trajectory data.

Motion Tracking

We analyzed the movement of most of our participants as they were approaching their choices. Due to camera recording failures, a subset of participants' trials were not successfully tracked. We present results from 24 (213 trials) dogs and 23 toddlers (205 trials) below.

As participants approached their first choice from the start box, we measured the maximum deviation from the optimal path to their first choice, regardless of whether that first choice was correct. We defined the optimal path as the straight line from the start position to the relevant cup (Figure 3.5). To evaluate optimal path deviation, we only considered the participant path data as

they were initially approaching their first choice, defined as coming within 0.3 meters of their chosen target. We then could compare this approach pattern for participants whose first choice was correct vs. incorrect. This approach allowed us to compare against a fixed optimal path across all participants, versus having to dynamically calculate the optimal way to approach the true target from the participants' placement in space and then their deviation from that line. This approach would have been particularly challenging for toddlers who sometimes went behind the pointer, or returned to the start box between trials.

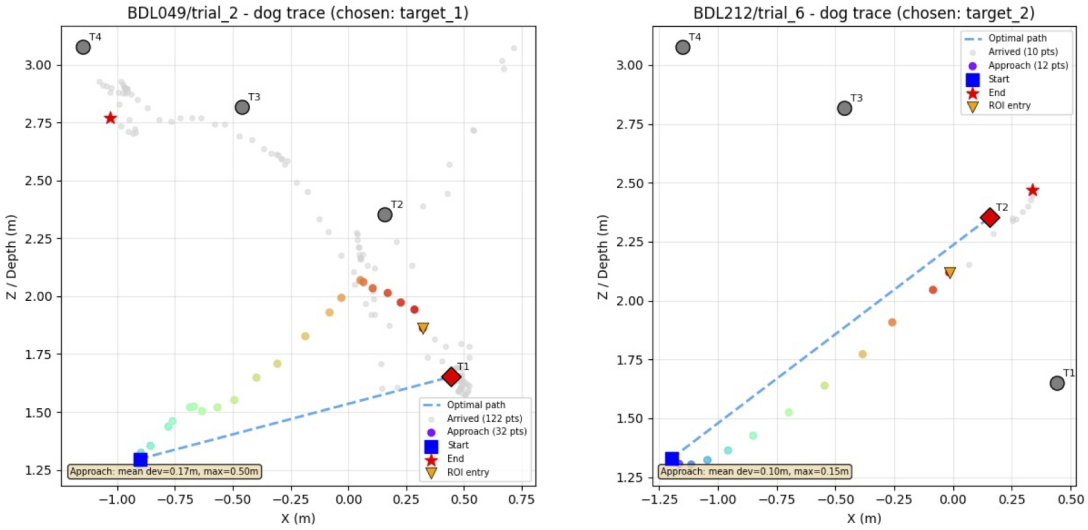

Figure 3.5: Example of paths taken by two dogs on two different trials. **Left:** This dog first chose cup 1 (outer-left) incorrectly and had a relatively large deviation from the optimal path. **Right:** This dog first chose cup 2 (inner-left) correctly and was close to the optimal path.

We examined the magnitude of deviation from the optimal path in two ways. The first considered how far the participant ever was from the optimal path they could have taken during their initial approach phase. Put another way, what is the maximum distance they were from the optimal path to their first chosen target. See Table 3.1

Grouping by their first chosen location, toddlers had the highest average maximum path deviation for location 4 (outer-right) (M = 0.588, SD = 0.37, Range = 0.05 - 1.82 m). In contrast, the shortest maximum path deviation was for approaches to location 1 (outer-left) (M = 0.435, SD = 0.41, Range = 0.05 - 1.69 m). When toddlers' first choice was the correct one, we see

consistently smaller deviations from the optimal path, with the exception of approaches to location 4, which had slightly more deviation from the optimal path when it was the true target.

For dogs, we see a different pattern. Like toddlers, dogs had the highest average maximum path deviation for location 4 (outer-right) (M = 0.438, SD = 0.34, Range = 0.03 - 1.30 m). Unlike toddlers, dogs had the shortest average maximum path deviation for location 3 (inner-right) (M = 0.389, SD = 0.33, Range = 0.00 - 1.38 m). When their first choice was correct, we consistently see smaller maximum path deviations than when their first choice was incorrect for the two inner locations. This pattern is reversed for the two outer locations. Specifically, when successful on their first search to one of the two inner cups, dogs tended to follow the optimal path more closely than when that first choice was incorrect. In contrast, when successful on their first search to one of the two outer locations, dogs tended to veer farther off the optimal path than when that choice was incorrect.

Table 3.1: Participant Maximum and Average Path Deviation by Location

	Maximum Deviation		**Average Deviation**	
Loc.	Mean (*SD*)	Range [Min, Max]	Mean (*SD*)	Range [Min, Max]
Toddlers				
1	0.435 (*0.407*)	[0.046, 1.692]	0.187 (*0.220*)	[0.019, 0.925]
2	0.565 (*0.381*)	[0.093, 1.464]	0.249 (*0.232*)	[0.026, 1.189]
3	0.578 (*0.514*)	[0.000, 2.374]	0.228 (*0.259*)	[0.000, 1.197]
4	0.588 (*0.375*)	[0.048, 1.822]	0.235 (*0.203*)	[0.016, 1.061]
Dogs				
1	0.401 (*0.381*)	[0.000, 1.616]	0.178 (*0.195*)	[0.000, 1.000]
2	0.390 (*0.319*)	[0.029, 1.176]	0.163 (*0.132*)	[0.011, 0.596]
3	0.389 (*0.330*)	[0.000, 1.377]	0.168 (*0.146*)	[0.000, 0.711]
4	0.438 (*0.344*)	[0.030, 1.298]	0.182 (*0.157*)	[0.016, 0.574]

We also evaluated deviation from the optimal path as an average over the course of the approach. Unlike with the maximum path deviation, the average optimal path deviation for toddlers was the largest for location 2 (inner-left)(M = 0.249, SD = 0.23, Range = 0.03 - 1.19 m). The smallest average path deviation was for approaches to location 1 (outer-left) (M = 0.187, SD = 0.22, Range = 0.02 - 0.92 m), just as we saw for the maximum path deviation. We also

found that average path deviation was generally smaller when the first chosen cup was also the correct one, with the exception of location 3, where that pattern was reversed.

Dogs' average path deviation, like their maximum path deviation, was the largest for location 4 (outer-right) (M = 0.182, SD = 0.16, Range = 0.02 - 0.57 m). Their smallest average path deviation was for location 2 (inner-left) (M = 0.163, SD = 0.13, Range = 0.01 - 0.59 m), though location 3 (inner-right) was close (M = 0.168, SD = 0.15, Range = 0.00 - 0.71 m). Interestingly, we see a different pattern for average path deviation by location for dogs based on success. Unlike maximum path deviation, dogs that were correct in their choices to the right side of the experimenter's body (locations 3 and 4) tended to have higher average path deviations. Dogs that were correct to locations on the left side of the experimenter's body (locations 1 and 2) tended to have lower average path deviations.

When comparing the paths taken by both species, dogs and toddlers continue to approach the problem differently. As seen in the behavioral results, dogs behave differently between the inner and outer locations. They appear systematically biased towards the two locations closest to the experimenter. This is reflected in dogs smaller maximum and average deviations from the optimal path for the two cups closest to the pointer. This could be taken as a measure of confidence or speed of interpreting the gesture. When dogs approach to choose the inner locations, they are potentially more sure of that choice, taking a direct and efficient route. In contrast, when dogs chose one of the outer locations they wandered more, taking more curving paths and indirect approaches. This could be indicative of less confidence, or of simply trying to gather more information prior to making an actual choice.

Toddlers show smaller average path deviations when their first choices were to the outer locations (Table 3.1). This smoother approach pattern to the outer cups might suggest that the proximity of the experimenter is having some kind of distracting effect on the toddlers, however their choice data confirms they choose between the four cups at approximately equal rates. Toddlers approach patterns are much more variable than dogs, and they tended to deviate more from the most efficient route. This could reflect a challenge with focus or inhibitory control,

helping to explain the challenge toddlers had with the choice task. If toddlers are distracted or are losing focus during their approach they may wander more, taking less direct routes. It is also possible that toddlers are using an associative cue or some other lower-level mechanism, and they are approaching indirectly to acquire more information about the pointing gesture itself. Swooping around can help to acquire multiple angles and could be helping to resolve some ambiguity.

We analyzed the pointing gestures produced by our experimenters across trials for both species. In line with our procedure, we saw no instances of cross-body pointing, and experimenters equally used their left and right arms as they only pointed with the same-side arm. In contrast with prior work (Pelgrim et al., 2024) that sub-selected an "optimal" pointing frame for analysis, we opted to use the entirety of the produced pointing gestures. We felt this better reflected the experience of the point receiver, namely that the participant watches the point being produced and held, rather than just seeing the final position. See Table 3.2 for summaries of each of our potential pointing vectors.

Table 3.2: Performance from Experimenters Pointing for Dogs and Toddlers

	Distance(m)↓	Weighted Accuracy↑	Perplexity↓
Eye-Wrist	0.997 *(0.52)*	0.452 *(0.36)*	3.412 *(0.50)*
Nose-Wrist	1.006 *(0.49)*	0.454 *(0.36)*	3.434 *(0.48)*
Shoulder-Wrist	0.967 *(0.54)*	0.447 *(0.35)*	3.407 *(0.42)*
Elbow-Wrist	0.983 *(0.52)*	0.457 *(0.36)*	3.426 *(0.44)*

From our first evaluation metric, Euclidean accuracy, the best performing vector was the Shoulder-Wrist. In contrast, using weighted accuracy our best performing vector was the Elbow-Wrist. Finally, using perplexity the best performing vector was also the Shoulder-Wrist. These findings are different than in prior work where integration of the position of the face (through the Eye-Wrist or Nose-Wrist) performed best (Pelgrim et al., 2024).

When experimenters produced points, it appears that the best strategy to identify the referent would be to follow the arm only, ignoring the position of the face. This suggests that if a point follower is attempting to integrate information from the head and the gaze with the body

movements, they may perform worse. Interpretation of these results should be tempered, as the metrics are quite close across the vectors. However, this provides an interesting baseline for the kinds of points produced in an experimental context by a trained researcher, and in Chapter 4, we will compare these points to those produced by dog owners. See Chapter 6 for further discussion of pointing candidate vectors.

3.3 Experiment 2: Adults

Given the challenges observed by both dogs and toddlers on our complex search task, we wanted to establish a baseline for adult performance. We conducted an online study with an adapted version of our methods where participants had to locate a hidden object after watching a video of an experimenter pointing. This study was pre-registered.

3.3.1 Participants

The participants were 26 adults (13 = Male, 13 = Female, Mean Age = 38.81 years, Age Range = 21 - 60 years) recruited from Prolific to complete an online survey. The survey was estimated to take 5 minutes and participants were compensated $1 (maintaining an average hourly rate of $12 per hour). An additional 1 participant was excluded for failing our comprehension check (detailed below in procedure).

3.3.2 Ethics Note

All ethics guidelines were followed, and study procedures were approved by the Brown Institutional Review Board, Protocol Number 2022003224.

3.3.3 Procedure

Participants began by completing a warm-up comprehension check. This allowed us to confirm that adults understood the setup of the overall task, and helped us to screen for potential

bots. During our comprehension check, participants watched a video of the experimenter placing the ball inside one of the four cups in full view. Participants were then asked where the experimenter had hidden the ball, and this was repeated for each of the four cups. If participants did not select all four locations correctly, they were excluded from the task.

After passing our comprehension check, participants advanced to our test trials. Participants were shown instructions, telling them that they would "...see a person point to where they hid the ball. Please indicate where the ball is hidden." Each test trial required the participant to begin by watching a video clip. The clip of the trials was identical to what dogs and toddlers saw in person. The experimenter began by showing the object they were going to hide, and then turned around and hid the object inside one of 4 cups out of sight. The experimenter then placed the cups out and pointed to the cup containing the hidden item. As with our earlier study, participants were able to search exhaustively. If a participant selected the incorrect location, they were shown a still image of the experimenter pointing and their previous choice was no longer able to be selected. See supplementary materials for examples. Participants continued this procedure until they selected the correct hiding spot for the object. This was repeated for a total of 10 trials.

3.3.4 Results

This task was pre-registered. As expected, adults were significantly above chance (0.25) following the experimenter's point from video, M = 0.97, 95% CI [0.92,1.02], $t(25) = 30.76, p <$ .001. Every single participant (26/26) followed the point at above chance levels. Only 3 of our participants ever made an error, and no participant required more than 2 choices to successfully identify the target. All this suggests that following the point, even in this complex four-object search task, is trivial for adults, whose performance is at ceiling levels.

3.4 General Discussion

In Experiment 1, we found that although dogs and toddlers are both able to solve a complex 4-cup object search task at above chance levels, toddlers significantly outperform dogs. We also found that this difference can be explained by the location of the experimenter. Specifically, dogs are highly biased by the proximity of the experimenter. This resulted in a preference for the two inner search locations in dogs alone. In contrast, toddlers do not appear to be impacted by the location of the experimenter, choosing equally across the left and right sides of the body. Toddlers make errors but in a less systematic fashion than dogs.

In Experiment 2, we demonstrated that adults are easily able to identify the target of the point and are nearly perfect in their overall performance. These findings are consistent with previous work showing that adults are generally skillful at comprehending pointing gestures (Cooney et al., 2018). Prior work has revealed that pointing comprehension is challenged in adults only when perspective failures occur (Herbort & Kunde, 2016; Herbort et al., 2021). Specifically, when pointers are actively producing points for a viewer, the viewer tends to make minor but systematic errors in identifying the true pointing target. This generally occurs due to a failure of the pointer to account for the different viewing angle of the point recipient. The fact that we did not see these kinds of errors in our sample is encouraging. It provides confidence that there was not a systematic production error on the part of the experimenter, or at least that the targets are sufficiently spaced to avoid ambiguity in an adult point viewer.

Dogs and toddlers both performed well below ceiling levels at this task. Particularly for toddlers, this finding challenges rich interpretations of point following. Infants begin engaging in joint attention, collaboratively engaging social partners by 12-15 months of age (Carpenter et al., 1995; Tomasello et al., 2005). The participants in this sample were typically developing toddlers between 17 and 24 months of age, long after joint attention abilities are established. There are multiple potential solutions for the discontinuity identified in this chapter.

First, the findings in this chapter do not suggest that toddlers of this age group cannot

engage in joint attention. There is a rich body of literature that provides great confidence in this ability. Rather, it suggests that toddlers may default to approaching pointing gestures using lower level cues. As mentioned above, there is no doubt that toddlers are capable of learning via associative mechanisms. Further, prior to acquiring full joint attention infants are able to respond appropriately to pointing gestures, following the point to the referent as young as 4 months of age (Bertenthal et al., 2014). Toddlers may default, in this relatively complex search task, to rely on associative cues or other lower level learned point following mechanisms. This may result in more ambiguity about the target of the referent or the task at hand. Toddlers may be able to localize the referent of the point and engage in joint attention, but this may be highly effortful. When the stakes are low, and they are able to search exhaustively, they may choose to engage in the easier mental approach, saving themselves the added cognitive control and load.

The low stakes of the task could also have contributed to toddlers (and potentially dogs) interpreting the pointing gesture as an informative one, but not as an imperative. Toddlers and dogs may have opted to, on some trials, prioritize independent exploration. Alternatively, they may have understood the general direction of the point but opted to investigate the other locations for the sake of completeness or to prolong the searching game. Across trials participants were highly motivated to locate the hidden item, but it is possible that the searching portion of the game was also rewarding.

It is not likely that this task presented a challenge to the working memory of the toddlers, as the pointer maintained their gesture until the participant located the item. This meant that toddlers could continually remind themselves of the social cue being provided. It is also possible that toddlers understand pointing as directing joint attention, but have inhibitory control challenges that prevent them from making optimal choices. Reflected in their highly variable approach paths, toddlers may become distracted between looking to the pointer and actually making a selection. If true, toddlers understand the "what" and the "why" of the pointing gesture, but when actually required to physically search the space they fail to inhibit incorrect motor choices.

In addition to the underlying cognitive processes, toddlers and dogs may also have difficulty actually identifying the referent of the pointing gesture. As mentioned above, adults tend to make minor but systematic errors when identifying pointing referents. This issue was not present in our adults, but it is possible that the margin of error is larger for toddlers and dogs. If the localization error is sufficiently large, participants could choose between items at random (toddlers) or based on alternative choice rule (closest to the pointer).

In ongoing work that is outside the scope of this dissertation, we are adding head-mounted eye-tracking to both species using an identical methodology. This ongoing study will help to further reveal the differences in the species, as well as examine if toddlers are displaying the key features of visual attention predicted for a higher-level cognitive account involving joint attention. Future research might also consider further varying or complicating the arrangement of the items presented. Particularly as the proximity of the pointer is a salient cue for dogs, items could be brought closer to the pointer. This may increase the ambiguity of the point itself, but could result in behavioral performance more reflective of dogs' interpretation of the gesture. Future work can also dive further into both the gesture and the path trajectory data presented here. Future work could consider integrating the path trajectory data with the pointing gesture data, considering the perspective of the participant as they moved through space. Particularly for toddlers on winding paths, pointing gestures may become increasingly ambiguous as they stray to farther sides of the testing space. Path trajectory could also be analyzed more dynamically, considering features like heading change or tortuosity over the course of trials.

Dogs are not the only species that are capable of following human pointing gestures (see review by Lea and Osthaus, 2018). Many other non-human animals are capable of responding to pointing gestures, particularly other domesticated species. Future work can continue to explore pointing comprehension through a comparative lens, examining how other species differ from human children.

These studies provide a promising first step to better understanding how pointing gestures are actually processed and the factors that might influence point following. They have shown that

toddlers do not, by default, follow points as levels expected by rich interpretations of pointing gestures in a 4-alternative search task. Further, they have shown that toddlers outperform dogs at search accuracy, and that this difference is explained by dogs' bias towards selecting the items closest to the experimenter. Dogs are generally more efficient in making their choices, possible reflecting less ambiguity in their decision making process or fewer challenges with inhibitory control or sustained focus.

References

Abidi, S., Williams, M., & Johnston, B. (2013). Human pointing as a robot directive. *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 67–68. https://doi.org/10.1109/HRI.2013.6483504

Agnetta, B., Hare, B., & Tomasello, M. (2000). Cues to food location that domestic dogs (*Canis familiaris*) of different ages do and do not use. *Animal Cognition*, *3*(2), 107–112. https://doi.org/10.1007/s100710000070

Aureli, T., Perucchini, P., & Genco, M. (2009). Children's understanding of communicative intentions in the middle of the second year of life. *Cognitive Development*, *24*(1), 1–12. https://doi.org/10.1016/j.cogdev.2008.07.003

Azari, B., Lim, A., & Vaughan, R. (2019). Commodifying Pointing in HRI: Simple and Fast Pointing Gesture Detection from RGB-D Images. *2019 16th Conference on Computer and Robot Vision (CRV)*, 174–180. https://doi.org/10.1109/CRV.2019.00031

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose tracking. https://doi.org/10.48550/arXiv.2006.10204

Begus, K., & Southgate, V. (2012). Infant pointing serves an interrogative function. *Developmental Science*, *15*(5), 611–617. https://doi.org/10.1111/j.1467-7687.2012.01160.x

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science*, *8*(6), 492–499. https://doi.org/10.1111/j.1467-7687.2005.00440.x

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, *50*(8), 2036–2048. https://doi.org/10.1037/a0037152

Birch, S. A. J., Vauthier, S. A., & Bloom, P. (2008). Three- and four-year-olds spontaneously use others' past performance to guide their learning. *Cognition*, *107*(3), 1018–1034. https://doi.org/10.1016/j.cognition.2007.12.008

Brosseau-Liard, P. E., & Birch, S. A. (2010). 'I bet you know more and are nicer too!': What children infer from others' accuracy. *Developmental Science*, *13*(5), 772–778. https://doi.org/10.1111/j.1467-7687.2009.00932.x

Butterworth, G. (2003). Pointing is the royal road to language for babies. In *Pointing: Where language, culture, and cognition meet.* (pp. 9–33). Lawrence Erlbaum Associates Publishers.

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K. V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., … Feichtenhofer, C. (2025). SAM 3: Segment Anything with Concepts. https://doi.org/10.48550/arXiv.2511.16719

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint Attention and Imitative Learning in Children, Chimpanzees and Enculturated Chimpanzees. *Social Development*, *4*(3), 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

Catala, A., Mang, B., Wallis, L., & Huber, L. (2017). Dogs demonstrate perspective taking based on geometrical gaze following in a Guesser–Knower task. *Animal Cognition*, *20*(4), 581–589. https://doi.org/10.1007/s10071-017-1082-x

Cirio, G., Olivier, A.-H., Marchal, M., & Pettré, J. (2013). Kinematic Evaluation of Virtual Walking Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 671–680. https://doi.org/10.1109/TVCG.2013.34

Cooney, S. M., Brady, N., & McKinney, A. (2018). Pointing perception is precise. *Cognition*, *177*, 226–233. https://doi.org/10.1016/j.cognition.2018.04.021

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental psychology*, *34*(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

Eatherington, C. J., Mongillo, P., Lõoke, M., & Marinelli, L. (2021). Dogs fail to recognize a human pointing gesture in two-dimensional depictions of motion cues. *Behavioural Processes*, *189*, 104425. https://doi.org/10.1016/j.beproc.2021.104425

Erb, C. D., Moher, J., Sobel, D. M., & Song, J.-H. (2016). Reach tracking reveals dissociable processes underlying cognitive control. *Cognition*, *152*, 114–126. https://doi.org/10.1016/j.cognition.2016.03.015

Fink, P. W., Foo, P. S., & Warren, W. H. (2007). Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception*, *4*(1), 1–18. https://doi.org/10.1145/1227134.1227136

Fisher, C. F., Byrne, M., & Johnston, A. M. (2025). Dogs (*Canis familiaris*) copy-all refine-later where children (*Homo sapiens*) overimitate. *Journal of Comparative Psychology*. https://doi.org/10.1037/com0000426

Hare, B., Brown, M., Williamson, C., & Tomasello, M. (2002). The Domestication of Social Cognition in Dogs. *Science*, *298*(5598), 1634. https://doi.org/10.1126/science.1072702

Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive Foundations of Learning from Testimony. *Annual Review of Psychology*, *69*(1), 251–273. https://doi.org/10.1146/annurev-psych-122216-011710

Herbort, O., Krause, L.-M., & Kunde, W. (2021). Perspective determines the production and interpretation of pointing gestures. *Psychonomic Bulletin & Review*, *28*(2), 641–648. https://doi.org/10.3758/s13423-020-01823-7

Herbort, O., & Kunde, W. (2016). Spatial (mis-)interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 78–89. https://doi.org/10.1037/xhp0000126

Hu, Z., Xu, Y., Lin, W., Wang, Z., & Sun, Z. (2022). Augmented Pointing Gesture Estimation for Human-Robot Interaction. *2022 International Conference on Robotics and Automation (ICRA)*, 6416–6422. https://doi.org/10.1109/ICRA46639.2022.9811617

Jaswal, V. K., Croft, A. C., Setia, A. R., & Cole, C. A. (2010). Young Children Have a Specific, Highly Robust Bias to Trust Testimony. *Psychological Science*, *21*(10), 1541–1547. https://doi.org/10.1177/0956797610383438

Jelinek, F., Mercer, R. L., Bahl, L. R., & Baker, J. K. (2005). Perplexity—a measure of the difficulty of speech recognition tasks. *The Journal of the Acoustical Society of America*, *62*(S1), S63. https://doi.org/10.1121/1.2016299

Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics

Koenig, M. A., & Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, *76*(6), 1261–1277. https://doi.org/10.1111/j.1467-8624.2005.00849.x

Koenig, M. A., & Jaswal, V. K. (2011). Characterizing Children's Expectations About Expertise and Incompetence: Halo or Pitchfork Effects? *Child Development*, *82*(5), 1634–1647. https://doi.org/10.1111/j.1467-8624.2011.01618.x

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as Epistemic Request: 12-month-olds Point to Receive New Information. *Infancy*, *19*(6), 543–557. https://doi.org/10.1111/infa.12060

Kundey, S. M. A., Los Reyes, A., Royer, E., Molina, S., Monnier, B., German, R., & Coshun, A. (2011). Reputation-like inference in domestic dogs (Canis familiaris). *Animal Cognition*, *14*(2), 291–302. https://doi.org/10.1007/s10071-010-0362-5

Lakatos, G., Gácsi, M., Topál, J., & Miklósi, Á. (2012). Comprehension and utilisation of pointing gestures and gazing in dog–human communication in relatively complex situations. *Animal Cognition*, *15*(2), 201–213. https://doi.org/10.1007/s10071-011-0446-x

Lakatos, G., Soproni, K., Dóka, A., & Miklósi, Á. (2009). A comparative approach to dogs' (*Canis familiaris*) and human infants' comprehension of various forms of pointing gestures. *Animal Cognition*, *12*(4), 621–631. https://doi.org/10.1007/s10071-009-0221-4

Lea, S. E. G., & Osthaus, B. (2018). In what sense are dogs special? Canine cognition in comparative context. *Learning & Behavior*, *46*(4), 335–363. https://doi.org/10.3758/s13420-018-0349-7

Lind, J., Ghirlanda, S., & Enquist, M. (2019). Social learning through associative processes: A computational theory. *Royal Society Open Science*, *6*(3), 181777. https://doi.org/10.1098/rsos.181777

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Lyn, H., & Christopher, J. (2020). A point is not a point is not a point: Reinterpreting three basic kinds of point comprehension. *The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)*, 260–263. https://doi.org/10.17617/2.3190925

Maginnity, M. E., & Grace, R. C. (2014). Visual perspective taking by dogs (*Canis familiaris*) in a Guesser-Knower task: Evidence for a canine theory of mind? *Animal Cognition*, *17*(6), 1375–1392. https://doi.org/10.1007/s10071-014-0773-9

ManyDogs, Espinosa, J., Stevens, J. R., Alberghina, D., Barela, J., Bogese, M., Bray, E., Buchsbaum, D., Byosiere, S.-E., Cavalli, C., Dror, S., Fitzpatrick, H., Freeman, M. S., Frinton, S., Gnanadesikan, G. E., Guran, C.-N. A., Glover, M., Hare, B., Hare, E., … Walsh, C. (2023). ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension. *Animal Behavior and Cognition*, *10*(3), 232–286. https://doi.org/https://doi.org/10.26451/abc.10.03.03.2023

Marshall-Pescini, S., Colombo, E., Passalacqua, C., Merola, I., & Prato-Previde, E. (2013). Gaze alternation in dogs and toddlers in an unsolvable task: Evidence of an audience effect. *Animal Cognition*, *16*(6), 933–943. https://doi.org/10.1007/s10071-013-0627-x

Marshall-Pescini, S., Passalacqua, C., Ferrario, A., Valsecchi, P., & Prato-Previde, E. (2011). Social eavesdropping in the domestic dog. *Animal Behaviour*, *81*(6), 1177–1183. https://doi.org/10.1016/j.anbehav.2011.02.029

Mills, C. (2013). Knowing When to Doubt: Developing a Critical Stance When Learning From Others. *Developmental Psychology*, *49*(3), 404–418. https://doi.org/10.1037/a0029500

Nitzschner, M., Kaminski, J., Melis, A., & Tomasello, M. (2014). Side matters: Potential mechanisms underlying dogs' performance in a social eavesdropping paradigm. *Animal Behaviour*, *90*, 263–271. https://doi.org/10.1016/j.anbehav.2014.01.035

Pasquini, E. S., Corriveau, K. H., Koenig, M., & Harris, P. L. (2007). Preschoolers Monitor the Relative Accuracy of Informants. *Developmental Psychology*, *43*(5), 1216–1226. https://doi.org/10.1037/0012-1649.43.5.1216

Pelgrim, M. H., Espinosa, J., Tecwyn, E. C., Marton, S. M., Johnston, A., & Buchsbaum, D. (2021). What's the point? Domestic dogs' sensitivity to the accuracy of human informants. *Animal Cognition*, *24*(2), 281–297. https://doi.org/10.1007/s10071-021-01493-5

Pelgrim, M. H., He, I. X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024). Find it like a dog: Using Gesture to Improve Object Search. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*. https://escholarship.org/uc/item/0nk6w9fd

Pfandler, E., Lakatos, G., & Miklósi, Á. (2013). Eighteen-month-old human infants show intensive development in comprehension of different types of pointing gestures. *Animal Cognition*, *16*(5), 711–719. https://doi.org/10.1007/s10071-013-0606-2

Scott, J. P. (1958). Critical Periods in the Development of Social Behavior in Puppies: *Psychosomatic Medicine*, *20*(1), 42–54. https://doi.org/10.1097/00006842-195801000-00005

Sobel, D. M., & Corriveau, K. H. (2010). Children Monitor Individuals' Expertise for Word Learning. *Child Development*, *81*(2), 669–679. https://doi.org/10.1111/j.1467-8624.2009.01422.x

Sobel, D. M., & Finiasz, Z. (2020). How Children Learn From Others: An Analysis of Selective Word Learning. *Child Development*, *91*(6), e1134–e1161. https://doi.org/10.1111/cdev.13415

Sobel, D. M., & Kushnir, T. (2013). Knowledge matters: How children evaluate the reliability of testimony as a process of rational inference. *Psychological Review*, *120*(4), 779–797. https://doi.org/10.1037/a0034191

Song, J.-H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks. *Trends in Cognitive Sciences*, *13*(8), 360–366. https://doi.org/10.1016/j.tics.2009.04.009

Southgate, V., Maanen, C. V., & Csibra, G. (2007). Infant Pointing: Communication to Cooperate or Communication to Learn? *Child Development*, *78*(3), 735–740. https://doi.org/https://doi.org/10.1111/j.1467-8624.2007.01028.x

Taylor, J. L., & McCloskey, D. I. (1988). Pointing. *Behavioural Brain Research*, *29*(1), 1–5. https://doi.org/10.1016/0166-4328(88)90046-0

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Trenkle, A., Göhl, M., & Furmans, K. (2015). Interpretation of pointing gestures for the gesture controlled transportation robot "FiFi". *2015 Annual IEEE Systems Conference (SysCon) Proceedings*, 721–726. https://doi.org/10.1109/SYSCON.2015.7116836

Virányi, Z., Gácsi, M., Kubinyi, E., Topál, J., Belényi, B., Ujfalussy, D., & Miklósi, Á. (2008). Comprehension of human pointing gestures in young human-reared wolves (*Canis lupus*) and dogs (*Canis familiaris*). *Animal Cognition*, *11*(3), 373–387. https://doi.org/10.1007/s10071-007-0127-y

Wang, Y., Deng, J., Sun, A., & Meng, X. (2023). Perplexity from PLM Is Unreliable for Evaluating Text Quality. https://doi.org/10.48550/arXiv.2210.05892

Wittgenstein, L. (1989). *Philosophical investigations* (G. E. M. Anscombe, Trans.; 3nd ed., repr). Blackwell.

Zaine, I., Domeniconi, C., & Wynne, C. D. L. (2015). The ontogeny of human point following in dogs: When younger dogs outperform older. *Behavioural Processes*, *119*, 76–85. https://doi.org/10.1016/j.beproc.2015.07.004

CHAPTER 4

Components of Naturalistic Point Production by Dog Owners [1]

4.1 Background

Pointing is a distinct social-communicative gesture that can have a variety of meanings based on context. A pointer may use the gesture to request an item (i.e., bring me the cup), provide information (i.e., your cup is right there), or begin a shared experience (i.e., let us look at that cool cup together) (Tomasello et al., 2007). When pointing, a person may use their head, body, hand, and arm to refer to an object or location in the environment. Pointing is intuitive for people, but what exactly are we doing when we point? And how does this sequence of unified actions help the viewer of the point to identify our shared target?

4.1.1 Development of Point Production

Point production develops early in life. Infants begin forming the classic hand shape of pointing (fist with the index finger extended) when they are 3 months old (Hannan & Fogel,

[1]This chapter is an expansion and extension of Pelgrim et al. (2024) published in the Proceedings of the Cognitive Science Society

1987). Infants may point with varying hand positions. Pointing in young children sometimes occurs with a flat hand (hand and arm outstretched with relatively straight wrist). However, a classic index finger shape is typically the canonical point. Interestingly, points are not used in a communicative fashion until much later in the first year of life, emerging between 9-12 months (Carpenter et al., 1998). Index finger pointing for social purposes emerges in infants by 12 months. Remarkably, this is true across cultures, observed in naturalistic parent-child interactions in locations ranging from rural Papua New Guinea to urban Japan (Liszkowski et al., 2012).

Even after pointing is social, point production can serve asocial, or private, purposes. Often compared to private speech, human beings of all ages point for themselves to facilitate understanding and memory. When trying to remember the location of a hidden toy, toddlers between the ages of 2 and 4 spontaneously produce index finger pointing gestures to help themselves remember (Delgado et al., 2011).

As discussed in Chapter 1, young children point for a variety of reasons. Proto-declarative pointing is pointing that serves to direct the attention of the viewer or provide information to that viewer. Proto-imperative pointing serves to instruct or command the viewer. Both pointing types are produced and understood by 12-month-old children (Bates et al., 1975; Tomasello et al., 2007).

Descriptions of points used in point following tasks are often qualitative, using general positional descriptions (i.e., pointing with the arm across the body) that lack standardization across studies. In young children, point production has been studied in the lab and in relatively naturalistic contexts, using toddlers' natural preferences and instincts. For example, to elicit declarative pointing gestures, experimenters had toys move in exciting ways while the infant could not reach the toy. The movement occurred behind the experimenters' back and the parent pretended not to notice. Infants are typically motivated to share interesting and exciting experiences with others. In this situation, infants spontaneously produce pointing gestures to engage their parent (Carpenter et al., 1998).

Ecological validity in infant pointing has also led to highly realistic tasks, including decorated room paradigms. In these, mothers and infants freely explore and observe a room filled with toys and decorations. Mothers are provided with minimal instructions, and the vast majority of infants spontaneously produce pointing gestures for their mothers while investigating the space (Liszkowski & Tomasello, 2011). In all we can have great confidence that human beings point socially early in development and do so readily across contexts.

In adults (when studied without children), research examining point production has often focused on tightly controlled settings. Adult humans point for each other at a specific item in a numerical or graphical array. Adults are generally quite accurate at producing points for others. However, there are minor but systemic errors in how the viewer perceives the targets of points (Herbort & Kunde, 2016; Herbort et al., 2021). Much of this has to do with errors in perspective taking, with researchers suggesting that the pointer fails to account for the different viewing angles of the viewer.

People likely point differently when they point for other adults in a tightly controlled setting versus another social partner, such as their dog, on a dynamic search task. Although there has been important work on how people understand and produce points when with their children, there has yet to be a systemic and quantitative investigation of adult naturalistic point production in other contexts. Gaze and body position are both salient cues that can provide information about the pointer's target, and we do not know how these sources are produced (and if they are easily separable). How do people naturally produce pointing gestures for their dogs, and is there are specific gestural components that may predict successful target localization?

4.1.2 Quantifying Pointing

Anecdotally, a common critique of point-following studies requiring dogs to search the space (like the one described in Chapter 3) is the artificial or stilted nature of the task. Requirements for tight experimental controls mean that experimenters may be inadvertently reducing the ecological validity of their studies. Are the points we are using similar to those that dogs see in

their real-world environments?

To better understand the kinds of gestures people produce, we can focus on quantifying body language cues. Inspired by prior work in robotics, we can assess points mathematically, generating lines (or vectors) from candidate body parts. In computer science and robotics, quantifying gestures is a necessity as following the points requires a programmed algorithm (Azari et al., 2019; Hu et al., 2022; Trenkle et al., 2015). In this study, we can take the reverse approach. Rather than trying to follow a point, we can see how people produce them. From each of the potential body language cues people produce, we can compare how closely they align with their true intended target. This provides insight into how pointers are approaching the problem.

As pointers naturally move their bodies, there are a number of potential cues that they can integrate or move separately. For example, qualitatively, a pointer may opt to point across their body with a straight elbow or a bent one. A pointer could opt to integrate their gaze with the movement of their hand, coordinating their looking patterns to coincide with hand movement. Quantifying these movements and gestures helps us to compare across pointers and avoid subjective coding judgments.

In past work, there are generally two categories of potential body cue lines. These lines or vectors are created by identifying two body parts of interest and drawing a line between them. That line is then projected into the world, and the intersection (or the distance of the intersection) of that line from the referent can be calculated.

The first category is based solely on the position of the arm itself. One could follow the line created by the forearm, tracing the elbow to the hand and continuing that line into space (Whitney et al., 2016). This approach (elbow-wrist vector) eliminates any consideration of body or head orientation. One could also follow the line from the shoulder to the hand. This approach (shoulder-wrist vector) integrates broad body angling, but does not account for bends in the elbow (Tölgyessy et al., 2017). The second category of pointing cues focuses on integrating

head orientation or gaze position with that of the hand. One could attempt to identify the target of gaze of the pointer and follow the line from the gaze to the hand. This approach (eye-wrist) would be an exceptionally poor cue if the pointer never looks towards the referent. Alternatively, one could omit following the gaze but rather follow broad head orientation, tracing a line from the nose to the hand. This approach (nose-wrist) likewise would be a poor cue if the pointer looks at an object that is not their pointed referent (Abidi et al., 2013; Azari et al., 2019).

If dog owners point for their dogs like trained experimenters, we would expect that their arm cues, as measured by the shoulder-wrist and elbow-wrist vectors, will best align with the true target of the search. In contrast, if dog owners naturally integrate their gaze as part of their intentionally produced cue, we would expect the eye-wrist or nose-wrist cues to perform best. We know from the rich pointing interpretations discussed in Chapter 3 that point following is often suggested as a bid to initiate joint attention. If dog owners are attempting to engage in joint attention with their dogs, we can expect them to integrate their gaze cues. Dogs' ability to respond to these cues, relative to how they followed cues from an experimenter in Chapter 3, can also help to inform the underlying cognitive mechanism behind point following in dogs.

We quantified how dog owners produced points for their dogs. Using recorded data from points produced during a search task, we analyzed how different gestural components (measured as vectors) aligned with the target of the search. In addition to recording dogs' choice behaviors, we integrated dogs' path trajectories as they made their choices. This provides us with insight into the underlying cognitive processes of the dog as they made their choices (Cirio et al., 2013; Fink et al., 2007). To assess this trajectory, we examined the paths taken by dogs on their approach, calculating their deviation from the most efficient route to their chosen target. This chapter helps to provide further insight into a critical gesture, pointing. By assessing how adults naturally produce points on search tasks we can begin to determine the potential cues available to dogs that could inform their learning and behavior.

4.2 Methods

4.2.1 Participants

The participants were 16 dog-guardian pairs. 16 dogs (7 = Female, 9 = Male, Mean Age = 63.56 Months, Age Range = 8 - 162 Months) and 16 adults (15 Women, 1 Man, Mean Age = 44.07 Years, Age Range = 27 - 67 Years, 14 White, 1 Native American or Alaskan Native, 1 Black or African American). An additional 2 dyads were excluded due to (1) Equipment failure resulting in data loss (N = 1), and (2) failure to complete all trials due to dog discomfort (N = 1)

4.2.2 Ethics Note

All ethics guidelines were followed, and study procedures were approved by the relevant ethics committees. All components of the study involving dogs were approved by the Brown Institutional Animal Care and Use Committee, Protocol Number 23-09-0002. Components of the study involving adult humans were approved by the Brown Institutional Review Board Protocol Number STUDY00000124.

4.2.3 Procedure

Throughout the task, dogs worked on leash and were handled by a trained experimenter, waiting in the start box approximately 2.25 m away from their owner. See Figure 3.1. Treats were hidden inside of silicone snack cups with integrated lids, which prevented the dog from visually inspecting the contents of the cup prior to making a choice. Choice was defined as physical contact with a cup, and all searches were exhaustive, meaning dogs could search cups until they located the one with the treat. If the dog had difficulty searching the cup the owner assisted them in getting the treat out.

Warm-Up Activities

Prior to starting the experiment, dogs engaged in a warm-up activity to receive a treat at each of the 4 hidden search locations with baiting visible. This warm-up also served to familiarize the owner with the task procedure. Dog owners received verbal instructions prior to beginning the task, and throughout trials as needed. Dogs watched the owner place a cup onto the ground. The owner then showed the dog a treat and placed the treat into a second cup in their hand. This cup was placed onto the opposite location to where the first cup already was (i.e., if the first cup was at the far right location the second cup would be placed on the far left). Dogs were then released to search for the treat. This was repeated for 4 trials so the dog received a treat from each cup at each search location, and owners were comfortable with general task procedures.

Test Trials

After completing warm-up trials, the experimenter gave the dog owner verbal instructions on how to complete the test trials. A trial began when the owner showed the treat. The owner then turned their back and hid the treat inside one of the four silicone lidded cups on a small table behind them. The owner then placed the cups on the search locations. The location of the hidden treat was semi-randomized so that dogs saw a total of 6 trials to the two locations closest to their owner and 6 to the locations farthest from their owner. The treat was also hidden an equal number of times on the left and right sides of the owner. Search locations were numbered 1-4, starting with the owner's outer left side (cup 1). Cup 2 was on the inner left, cup 3 was on the owner's inner right, and cup 4 was on the far right.

Owners were instructed where to place the cup containing the treat at the start of each trial. After placement, the owner used whatever words and pointing gestures "felt natural" to instruct their dog on the location of the hidden item. When they felt ready for their dog to approach they gave a verbal release word (e.g., "Okay", "Free") which cued the experimenter to release the dog to make their approach. Dogs were allowed to search exhaustively, meaning they could explore each location until the treat was found. After locating the treat, dogs were allowed to eat

the treat and were then recalled by the experimenter. This procedure was repeated for a total of 12 trials.

Across the test trials, we recorded both the owner's body language and the dogs' movement in space. We used two Intel RealSense cameras to record both depth and RGB data. This allowed us to track the experimenter's body positions and quantify their pointing, helping to better understand what kind of gestural cues were available to participants (Pelgrim et al., 2024). We also tracked the participants' movements during their choices, capturing their position and speed as they approached.

Exploratory Trials: Pointing for an Experimenter

As an exploratory measure, when possible we also had dog owners provide points for a human experimenter. Unlike in warm-up and test trials, the dog was not involved in this phase of the study. A second experimenter held the dog's leash to prevent them from approaching during this section. An experimenter sat in a chair behind the dog start box and asked the owner to point at a cup for them. The owner was instructed to return to rest in between each point but did not receive any constraints on the timing of their points or of their pointing style. The sequence was identical to the one used by their dog (12 trials to the same locations), but this task did not involve any searching.

Given the requirement for the dog to be distracted by an experimenter, not every dog owner was able to complete all 12 exploratory trials. If the dog became distressed or was unwilling to stay with the experimenter the control trials were stopped. All but one participant (15/16) were able to complete some of the possible 12 exploratory trials.

4.3 Data Coding & Preparation

4.3.1 Dog Behavioral Data

During the session, all of dogs' choices were coded live by a research assistant. A subset of dogs (25% or 4/16 dogs) were re-coded from video by a research assistant not previously involved in the sessions. Coders agreed on all trials (48/48 or 100%).

4.3.2 Motion Tracking

We used Google's MediaPipe Pose Landmarker (Bazarevsky et al., 2020) to process input RGB images and detect key points on the human body. We then employed the depth information to transform the relevant key points' coordinates from 2D into 3D space. As in past work inspired by robotics and gesture tracking more broadly, we then converted the experimenter's body position into a projected vector in the world (Pelgrim et al., 2024).

Unlike in Chapter 3 where we could pre-specify the pointing arm used, we had to identify the pointing arm used by owners to project our vectors. Initially, we attempted to run this automatically, using the arm that moved furthest from the body over the course of the trial. Unfortunately, this method was built on the assumption of lateral or same-side arm pointing, and did not work when the person pointed across their body (as the pointing arm was exceptionally close to their center of mass). With this limitation in mind, we manually reviewed all videos to identify if the owner used their left or right arm. Once the person and their true gesture (via manual selection of the pointing arm) was represented in three dimensions, we ray-casted our candidate vectors and calculated the vectors' intersection points with the cups.

We initially examined five potential vectors that have shown promise in robotics. It was possible our participants were using one or a combination of these, and we present three metrics of the accuracy of these candidate vectors in our results section. Details of the metrics and their precise calculations can be found in Chapter 3. Our five candidate pointing vectors were the Eye-Wrist, Nose-Wrist, Shoulder-Wrist (whole arm), Elbow-Wrist (forearm), and Gaze.

As in Chapter 3, we had difficulty tracking and producing the gaze vector from our pointer. This was primarily observed as dog owners regularly bent down and angled their heads and bodies towards the cups such that at least one of their key facial features was obscured. Given this inaccuracy, and the fact that the gaze vector has been demonstrated to perform poorly as a localizer of a pointing target in naturalistic point production, we opted to proceed without the gaze vector (Pelgrim et al., 2024).

Dog Path

In addition to tracking the human pointing gestures, we also evaluated the position of the dogs over the course of trials. We captured how efficiently or directly dogs moved towards their first choice. Dogs might move along a more direct path when they are confident about the location of the hidden item. However, if dogs are more confused or are waiting to select between two items they might take a less efficient path towards their eventual choice. For each frame from our videos, we identified the dog using Segment Anything Model 3 (SAM 3) (Carion et al., 2025). We used a bound-box approach to identify the location of the dog without attending to skeletal tracking as that posed challenges during the dog's position changes and movement. We then integrated the position of the dog with the depth data, creating a trace of the dog's path across frames (Figure 4.1).

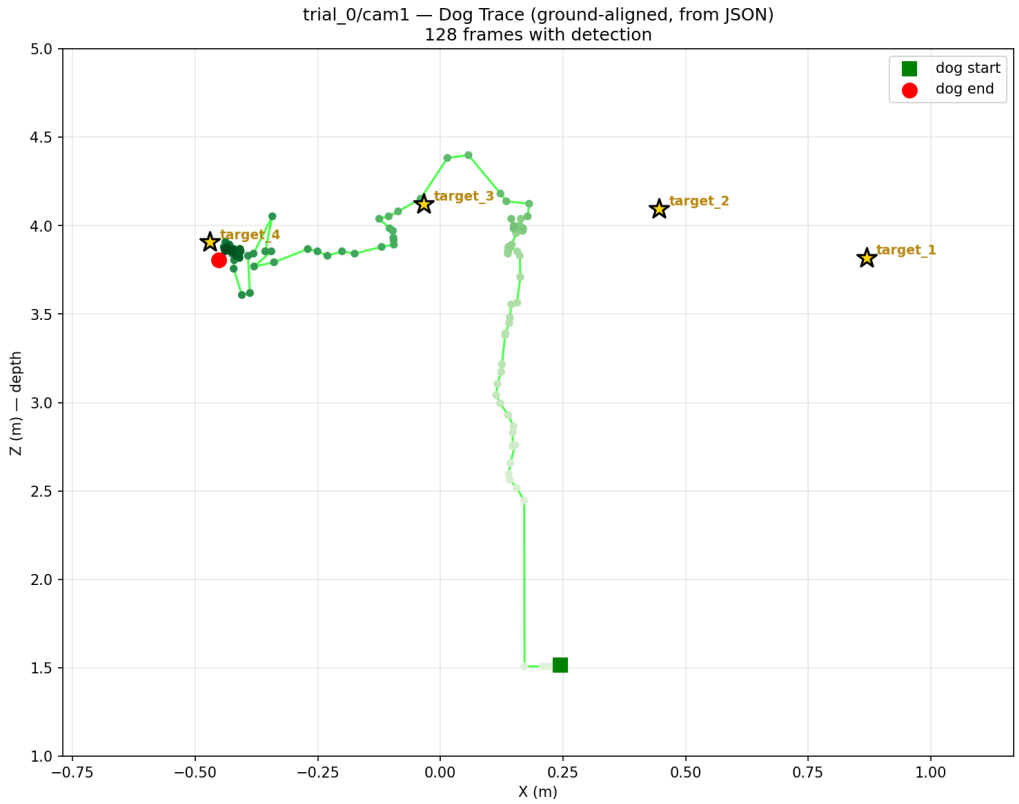

Figure 4.1: Trace of the detected path of a dog on a single trial. The treat is in Target 4.

4.4 Results

4.4.1 Dog Behavioral Data

Dogs were generally able to follow points from their owners to find the hidden treat on our 4-cup search task. We conducted a one-sample t-test against chance levels (1/4 or 0.25). Dogs located the treat on the first try at levels significantly above chance, M = .41, 95% CI [.33, .48], $t(15) = 4.61, p < .001$. Of our 16 dogs, 13 (81.3%) located the treat on the first try at above chance levels (Figure 4.2).

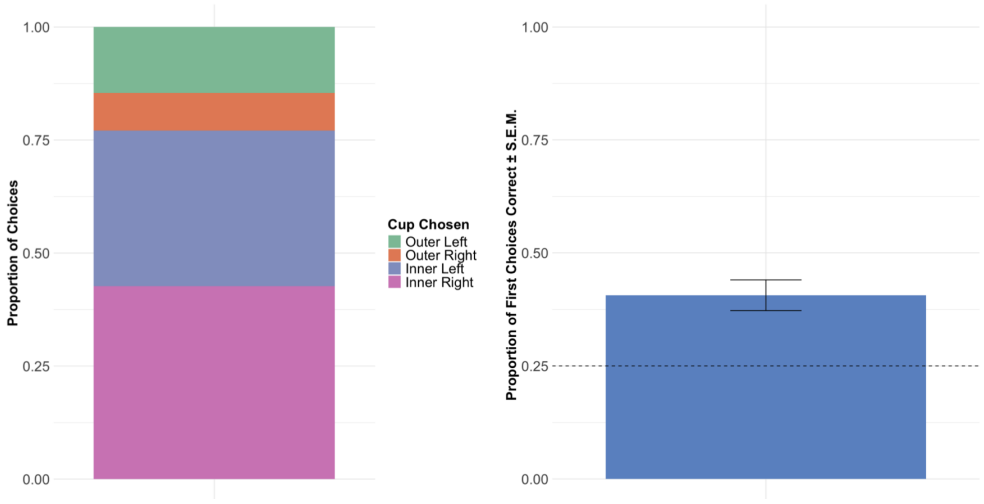

Figure 4.2: **Left:** Proportion of trials dogs chose each of the four cups first. **Right:** Proportion of first choices correct

Dogs appear to be attracted to their owner, choosing the inner locations (meaning those closer to their owner) 77% of the time. We examined dogs' choices based on whether they first chose the cup containing the item on the correct side of their owner (the Left or Right) rather than the actual item itself. When averaging by the side of the owner, dogs performed above chance (50%) levels, and chose the correct side of the room 72.9% of the time. This suggests that dogs may be interpreting their owners' gestures as broadly indicating a room side, rather than a specific referent. Alternatively, it is possible that dogs were biased to approach their owner and up-weight the proximity of the human above the gestures they are producing. Finally, the gestures produced by dog owners may be increasingly vague when indicating more distant referents, leading to increased confusion by the dogs. To better understand these possibilities, we looked to the gestures produced by dog owners.

4.4.2 Owner Point Production

We analyzed the pointing gestures produced by dog owners across trials. Participants varied widely in their choice of pointing style. Participants generally preferred pointing with their right arm. Across our trials, participants used their right arm to point for their dog on 117/192

Table 4.1: Performance from Humans Pointing for their Dogs

	Distance(m)↓	Weighted Accuracy↑	Perplexity↓
Eye-Wrist	0.535 *(0.26)*	0.685 *(0.35)*	3.432 *(0.46)*
Nose-Wrist	0.554 *(0.26)*	0.679 *(0.35)*	3.445 *(0.43)*
Shoulder-Wrist	0.604 *(0.23)*	0.670 *(0.35)*	3.60 *(0.41)*
Elbow-Wrist	0.751 *(0.26)*	0.657 *(0.35)*	3.722 *(0.31)*

(61%). On the majority of our trials, participants used the same-side arm to point (155/192). However, participants did still point across their bodies some of the time. The majority (12/16 or 75%) of pointers used their opposite arm (or pointed across their body) on at least one trial. Three of our pointers pointed across their body on half of more of trials, but no pointer used a cross-body gesture on every trial, suggesting that this is a pointing style that only feels useful for dog owners some of the time. In comparison to Chapter 3, we see that owners naturally point in a more variable way than experimenters. This is not surprising as owners were given no instructions or constraints about what arms to use, though it is interesting that this variance in arm placement and usage does not seem to be impacting dogs' choice behaviors.

Summary metrics for each of our candidate vectors can be found in Table 4.1. As a first metric, we considered the average Euclidean distance from the ground intersection of our candidate vectors to the true target. In contrast with prior work (Pelgrim et al., 2024) that sub-selected an "optimal" pointing frame for analysis, we opted to use the entirety of the produced pointing gestures. We felt this better reflected the experience of the point receiver, namely that the participant watches the point being produced and held, rather than just seeing the final position. Considering the average performance of our candidate vectors across the trial, the eye-to-wrist vector had the lowest average distance to true target in meters (M = 0.535, SD = 0.26).

As a secondary evaluation metric, we examined the weighted accuracy of the vectors. As with Euclidean distance, the best performing vector when using weighted accuracy was also the eye-to-wrist vector. The eye-to-wrist vector had the highest weighted accuracy (M = 0.685, SD = 0.35). Finally, in keeping with the trend in our other two metrics the best performing vector

(meaning lowest in perplexity) was the eye-to-wrist vector (M = 3.432, SD = 0.46). However, the nose-wrist vector was comparable as captured by perplexity (M = 3.445, SD = 0.43). Taken together, all three of our evaluation metrics suggest that the best body cue to follow from the naturally produced points involves integrating cues from the gaze with those from the end of the hand.

As an exploratory measure, we also examined the distances to targets by vector as a function of the target number. It was possible that pointers were more accurate, as measured by Euclidean distance, for closer targets (for example) than for further targets, regardless of vector. It was also possible that certain vectors underperformed at further locations than at closer ones. We conducted a mixed-effects linear regression predicting Euclidean distance to the true target as a function of the location of the target (cup 1, 2, 3, or 4), the vector type, and a random intercept per participant. For the purpose of this analysis, we averaged across pointed frames to calculate the Euclidean distance averaged for each trial. If we keep each of the individual frames, the results do not change significantly.

We found that distance to the target changed as a function of the location of the target, $F(3, 741.01) = 19.008$, $p < .001$. Distances were generally shorter (indicating greater accuracy) for the search targets on the right side of the owner's body than on the left. Although we do not have data on the handedness of our participants, this may be a result of the right-hand dominance of most people. Averaging across vector types, the average Euclidean distance to the true target was smallest for target 4 (M = 0.540, SD = 0.30) and was highest for target 2 (M = 0.697, SD = 0.25). This is indicative of a broader pattern where owners were more accurate (indicated by a smaller Euclidean distance) to the outer locations than to the inner locations. Post-hoc pairwise comparisons with a Tukey correction revealed that accuracy was significantly different ($p < .05$) between the average Euclidean distance of pointing vectors between location 1 vs. 2, 1 vs. 4, 2 vs. 3, and 2 vs. 4. We saw no statistically significant difference between locations 3 vs. 4. Notably, the closest distances from owners were for those that their dogs' did not tend to choose, the outer locations. It is possible that dog owners were aware of their dogs' proximity bias and

were more careful in their gestures for farther locations.

As discussed previously (see Table 4.1) when examining Euclidean distances to the true target the Eye-Wrist had the lowest average distance and the Elbow-Wrist had the highest (meaning performed the worst). From our model, we found a significant effect of vector type on distance to target, $F(3, 740.97) = 41.509$, $p < .001$. Post-hoc comparisons with a Tukey correction found a statistically significant difference ($p < .05$) between the elbow-to-wrist vector (our farthest) and our other three candidate vectors. We also found a statistically significant difference between the eye-to-wrist vector and the shoulder-to-wrist vector. Finally, we found little variance explained by participant ($ICC = 0.243$).

Given that pointing gestures were in fact different as a function of the target, we examined whether dogs' performance related to the Euclidean distance of our best performing candidate vector, the eye-to-wrist vector. We conducted a mixed-effects logistic regression predicting whether dogs' first choice was correct (measured as a binary variable) as a function of the Euclidean distance of the top performing vector to the true target (the eye-to-wrist vector) and the location of the target, with the random effect of participant. In keeping with findings reported earlier, we saw a significant effect of location on success, χ^2 (3, $N = 191$) = 32.803, $p <.001$. Post-hoc pairwise comparisons with a Tukey correction found a statistically significant ($p < .05$) difference between dogs' choice of location 1 and the two inner locations, as well as location 4 and the two inner locations. There was not a statistically significant difference between the two inner locations (locations 2 and 3) or the two outer locations (locations 1 and 4).

We did not anticipate that there would be a statistically significant relation between Euclidean distance of the vector, given that the pattern of Euclidean distances we observed by location was the opposite of dogs' choice behaviors. Specifically, dog owners tended to have smaller distances (or more "accurate" points) to the farther locations, and have farther locations or more ambiguous points to closer locations. From the same model, we did not see a significant effect of the Euclidean distance of the individual point, χ^2 (1, $N = 191$) = 0.027, $p = .870$. This may indicate that the differences between the Euclidean distance from the vector to the locations,

while statistically significant, do not translate into perceptible differences on the part of the dog. It is also possible that our findings are biased by dogs' strong preference for the two inner locations. Dog owners' potential efforts to make precise and unambiguous gestures for the farther locations did not appear to translate into improved choice task success. We saw little variance explained by participants from the random effect of our model ($ICC = 0.010$).

Dog Path

In addition to tracking the gestures produced by owners, we evaluated the paths taken by dogs over the course of the trial. Due to depth artifacts present in one participant's video, 4 of their 12 trials had to be dropped, resulting in that dog contributing 8 trials to this section. We present results from 188 trials across our 16 dogs here. See Table 4.2 for summary statistics on path deviation.

Table 4.2: Dog Maximum and Average Path Deviation by Location

	Maximum Deviation		**Average Deviation**	
Loc.	Mean (*SD*)	Range [Min, Max]	Mean (*SD*)	Range [Min, Max]
1	0.399 (*0.328*)	[0.024, 1.480]	0.112 (*0.105*)	[0.007, 0.382]
2	0.255 (*0.182*)	[0.021, 0.855]	0.062 (*0.058*)	[0.005, 0.307]
3	0.319 (*0.313*)	[0.030, 1.232]	0.066 (*0.064*)	[0.006, 0.378]
4	0.406 (*0.237*)	[0.052, 0.970]	0.102 (*0.066*)	[0.013, 0.239]

We examined how close dogs were on their approach path to the most efficient route. We defined approach path as starting when the dog was waiting in their start box, just prior to being released to start their search. The approach ended once they were at their first choice, defined as within less than 0.3 m of the first chosen target. For the duration of this path, we compared the true path taken by the dog with the most efficient route they could have taken. We operationalized the most efficient route as the optimal path, defined as the straight line from the start position to the relevant item. We were most interested in comparing their approach to the optimal path for dogs' first choice

We examined deviation from the optimal path in two ways. The first was to consider how far,

during their initial approach phase, the participant ever was from the optimal path they could have taken. Put another way, what is the maximum distance they were from the optimal path to their first chosen target? Grouping by their first chosen location, dogs tended to deviate more from the optimal path when their first choice was one of the two outer locations than when it was one of the inner ones. Dogs had the highest average maximum path deviation for location 4 (M = 0.406, SD = 0.24, Range = 0.05 - 0.97 m). In contrast, the shortest maximum path deviation was for approaches to location 2 (M = 0.255, SD = 0.18, Range = 0.02 - 0.85 m). When their first choice was the correct one, we saw consistently smaller deviations from the optimal path across all four targets.

We also evaluated deviation from the optimal path as an average over the course of the approach. Similar to the maximum path deviation, dogs generally deviated less from the optimal path when choosing the inner locations first, as compared to when they chose the outer locations first. Unlike with the maximum path deviation, the average optimal path deviation for dogs was the largest for location 1 (M = 0.112, SD = 0.11, Range = 0.01 - 0.38 m). The smallest average path deviation was for approaches to location 2 (M = 0.062, SD = 0.06, Range = 0.004 - 0.31 m), just as we saw for the maximum path deviation. We also found that the average path deviation was consistently smaller when the first chosen cup was also the correct one.

Dogs' path deviation when following points from an owner and an experimenter (Chapter 3) followed a similar pattern (Figure 4.3). Dogs took the most direct route to the inner locations and took more indirect paths to choose the outer locations. When following points from an experimenter, dogs tended to deviate much more from the efficient route. This may be because there was less ambiguity when following gestures from their owners, or that dogs were more confident in their choices and needed less processing time. It is also possible that dogs were just generally more hesitant to approach the experimenter in Chapter 3, which could have resulted in hesitancy and path deviation as they resolved an internal conflict. However, this seems unlikely as dogs tended to first select the cups closest to the experimenter.

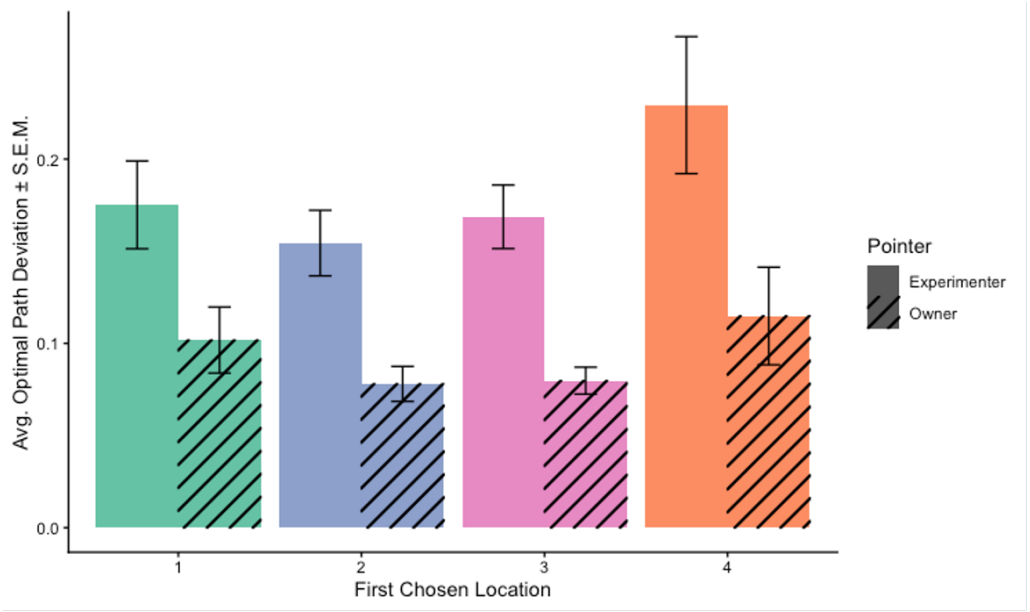

Figure 4.3: Average deviation from optimal path when following points from an owner and from an experimenter. Experimenter data is from Chapter 3.

Exploratory Phase: Pointing for an Experimenter

We analyzed the pointing gestures produced by dog owners for an experimenter after pointing for their dog. We collected a total of 167 trials. As mentioned above in the Data Coding & Preparation section, one dog owner was unable to contribute any data to this exploratory section due to dog stress. The minimum number of trials contributed in our sample was 5, and the modal contribution was 12.

We analyzed the points produced during this exploratory phase using the same three metrics as above, see Table 4.3. Consistent with points produced for their dog while the dog was actually searching, the most accurate vector as evaluated by Euclidean distance was the Eye-Wrist Vector (M = 0.464, SD = 0.24). Interestingly, when considering weighted accuracy, the shoulder-wrist vector performed the best (M = 0.670, SD = 0.35). This pattern differs from that seen in points produced for dogs while the dog is actually searching. Finally, the best performing vector as captured by perplexity was the Nose-Wrist (M = 3.450, SD = 0.39) however it was exceptionally close to the Eye-Wrist vector (M = 3.469, SD = 0.41).

Table 4.3: Performance from Human Pointing for an Experimenter

	Distance(m)↓	Weighted Accuracy↑	Perplexity↓
Eye-Wrist	0.464 *(0.24)*	0.636 *(0.36)*	3.469 *(0.41)*
Nose-Wrist	0.471 *(0.25)*	0.638 *(0.36)*	3.450 *(0.39)*
Shoulder-Wrist	0.638 *(0.34)*	0.670 *(0.35)*	3.711 *(0.37)*
Elbow-Wrist	0.785 *(0.29)*	0.618 *(0.34)*	3.840 *(0.21)*

We also examined the distances to targets produced for an experimenter by vector type as a function of the target number. We conducted a mixed-effects linear regression predicting Euclidean distance to the true target as a function of the location of the target (cup 1, 2, 3, or 4) the vector type, and a random intercept per participant. As with points produced for dogs, we found that distance to the target changed as a function of the location of the target, $F(3, 646.52) = 19.30$, $p < .001$.

Distances were generally shorter for the search targets on the right side of the experimenter's body than on the left. Like with points produced for their own dog, this may be influenced by the handedness of our participants. Averaging across vector types, the average Euclidean distance to true target was smallest for target 3 ($M = 0.517$, $SD = 0.30$) and was highest for target 2 ($M = 0.649$, $SD = 0.15$). Post-hoc pairwise comparisons with a Tukey correction revealed that accuracy was significantly different ($p < .05$) between the average Euclidean distance of pointing vectors between location 1 vs. 3, 1 vs. 4, 2 vs. 3, and 2 vs. 4.

From our model we also found a significant effect of vector type on distance to target, $F(3, 646.05) = 84.38$, $p < .001$. Post-hoc comparisons with a Tukey correction found a statistically significant difference ($p < .05$) between the elbow-to-wrist vector (our farthest) and our other three candidate vectors. We also found a statistically significant difference between the eye-to-wrist vector and the shoulder-to-wrist vector, as well as the nose-to-wrist vector and the shoulder-to-wrist vector. Finally, we saw a little variance explained by participant from the random effect in our model ($ICC = 0.213$)

4.5 Discussion

In this chapter, we have explored how owners freely point for their dogs and for an experimenter on an object search task. We found that dogs were successful on this task, performing significantly above levels expected by chance alone, but they are nowhere near ceiling. This was a challenging task for dogs, and they frequently made errors. Dogs preferentially selected the locations closest to their owners, first selecting the two inner locations on the vast majority of trials. This suggests that the proximity of their owner may be an overwhelmingly strong cue for dogs.

Our findings with dogs are highly consistent with those reported in Chapter 3 when dogs were following points produced by an experimenter. Specifically, dogs were generally following the point at levels significantly different from chance but not near ceiling. Further, they were generally skillful at going to the correct side of the room but struggle to avoid choosing the inner locations. It is interesting that dogs were behaving in a comparable way when the pointer was their owner and when it was a stranger (the experimenter). Further, owners often pointed in unique ways and we observed individual differences in factors like arm usage and cross-body pointing. However, this appeared to neither help nor hinder dogs' ability to find the hidden treat on this search task.

The challenge of approaching (or failing to avoid) their guardian was also reflected in the way that dogs approached to make their choices. Dogs were generally following the more efficient route when first selecting the inner cups, particularly when they were correct. In contrast, dogs deviated from the optimal route when selecting the outer cups. For the purposes of this exploration, we have only considered magnitude deviation from the optimal path. Future research might explore in depth what those deviations look like, and consider categorizing different approach paths based on location or based on dog success.

We applied four candidate vectors to naturalistically produced points and evaluated how closely those vectors aligned with the real-world pointing target. We found that when considering

Euclidean distance, weighted accuracy, and perplexity, the ray-cast vector from the eye to the wrist performed best. The close alignment of this particular vector with the true targets suggests that pointing gestures, as they are naturally produced, are best captured by integrating the positioning of the eyes (or the gaze) with the position of the wrist. This may mean that separating gaze from pointing severs the most naturally produced directions and may help to explain confusion seen in prior work. If a pointing gesture to a distant object is naturally produced by the pointer aligning their gaze with the movement of their hand, then forcing a pointer to gaze differently disrupts the cues they would naturally create. This may result in confusion on the part of the point viewer, who may be used to starting the processing of a distant point cue by attending to the position of the pointer's face. Future work can consider exploring what kinds of visual patterns point viewers use, potentially using wearable eye-tracking on a comparable search task.

When pointing for an experimenter, we saw some differences in the cue alignment relative to pointing for dogs. Anecdotally, many dog owners described feeling "a bit silly" and would spontaneously chuckle while pointing, as pointing for another adult human in such a constrained context may feel unnatural. Future work might consider exploring how people point for other non-adult agents, such as considering how points are naturalistically produced for a child or for a robot. Alternatively, the pointing task itself could be made more ambiguous and involve searching with a partner rather than the sort of instruction present in the current task.

The dataset we collected has a rich variety of data that may be explored by future researchers. As previously mentioned, pointing is generally described qualitatively in the literature. Future work might explore how naturalistic points produced by dog owners for their dogs would be described or categorized using a similar approach. Further, future analysis could consider coding for the vocalizations generated by dog guardians when pointing for their dogs. No vocalizations were produced when pointing for an experimenter, but people frequently called their dogs name, used attention-grabbing sounds, or provided verbal encouragement before searching began. This could be examined quantitatively to consider how people may be naturally incorporating cues

above and beyond body language to provide social-communicative information.

Finally, future work can consider expanding into other species. In piloting data, attempts to have parents naturalistically produce points for their toddlers failed due to the toddlers' unwillingness to wait away from the parent. Potentially by using two-parent households or by working with older ages, future researchers may apply the same approach we have outlined here to parents pointing for their children, and see how that differs from the human-dog context.

References

Abidi, S., Williams, M., & Johnston, B. (2013). Human pointing as a robot directive. *8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 67–68. https://doi.org/10.1109/HRI.2013.6483504

Azari, B., Lim, A., & Vaughan, R. (2019). Commodifying Pointing in HRI: Simple and Fast Pointing Gesture Detection from RGB-D Images. *2019 16th Conference on Computer and Robot Vision (CRV)*, 174–180. https://doi.org/10.1109/CRV.2019.00031

Bates, E., Camaioni, L., & Volterra, V. (1975). The Acquisition of Performatives Prior to Speech. *Merrill-Palmer Quarterly of Behavior and Development*, *21*(3), 205–226.

Bazarevsky, V., Grishchenko, I., Raveendran, K., Zhu, T., Zhang, F., & Grundmann, M. (2020). BlazePose: On-device Real-time Body Pose tracking. https://doi.org/10.48550/arXiv.2006.10204

Carion, N., Gustafson, L., Hu, Y.-T., Debnath, S., Hu, R., Suris, D., Ryali, C., Alwala, K. V., Khedr, H., Huang, A., Lei, J., Ma, T., Guo, B., Kalla, A., Marks, M., Greer, J., Wang, M., Sun, P., Rädle, R., … Feichtenhofer, C. (2025). SAM 3: Segment Anything with Concepts. https://doi.org/10.48550/arXiv.2511.16719

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age. *Monographs of the Society for Research in Child Development*, *63*(4), i–174. https://doi.org/10.2307/1166214

Cirio, G., Olivier, A.-H., Marchal, M., & Pettré, J. (2013). Kinematic Evaluation of Virtual Walking Trajectories. *IEEE Transactions on Visualization and Computer Graphics*, *19*(4), 671–680. https://doi.org/10.1109/TVCG.2013.34

Delgado, B., Gómez, J. C., & Sarriá, E. (2011). Pointing gestures as a cognitive tool in young children: Experimental evidence. *Journal of Experimental Child Psychology*, *110*(3), 299–312. https://doi.org/10.1016/j.jecp.2011.04.010

Fink, P. W., Foo, P. S., & Warren, W. H. (2007). Obstacle avoidance during walking in real and virtual environments. *ACM Transactions on Applied Perception*, *4*(1), 1–18. https://doi.org/10.1145/1227134.1227136

Hannan, T. E., & Fogel, A. (1987). A case-study assessment of "pointing" during the first three months of life. *Perceptual and Motor Skills*, *65*(1), 187–194. https://doi.org/10.2466/pms.1987.65.1.187

Herbort, O., Krause, L.-M., & Kunde, W. (2021). Perspective determines the production and interpretation of pointing gestures. *Psychonomic Bulletin & Review*, *28*(2), 641–648. https://doi.org/10.3758/s13423-020-01823-7

Herbort, O., & Kunde, W. (2016). Spatial (mis-)interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(1), 78–89. https://doi.org/10.1037/xhp0000126

Hu, Z., Xu, Y., Lin, W., Wang, Z., & Sun, Z. (2022). Augmented Pointing Gesture Estimation for Human-Robot Interaction. *2022 International Conference on Robotics and Automation (ICRA)*, 6416–6422. https://doi.org/10.1109/ICRA46639.2022.9811617

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A Prelinguistic Gestural Universal of Human Communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., & Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cognitive Development*, *26*(1), 16–29. https://doi.org/10.1016/j.cogdev.2010.10.001

Pelgrim, M. H., He, I. X., Lee, K., Pabari, F., Tellex, S., Nguyen, T., & Buchsbaum, D. (2024). Find it like a dog: Using Gesture to Improve Object Search. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*. https://escholarship.org/uc/item/0nk6w9fd

Tölgyessy, M., Dekan, M., Duchoň, F., Rodina, J., Hubinský, P., & Chovanec, L. (2017). Foundations of Visual Linear Human–Robot Interaction via Pointing Gesture Navigation. *International Journal of Social Robotics*, *9*, 1–15. https://doi.org/10.1007/s12369-017-0408-9

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Trenkle, A., Göhl, M., & Furmans, K. (2015). Interpretation of pointing gestures for the gesture controlled transportation robot "FiFi". *2015 Annual IEEE Systems Conference (SysCon) Proceedings*, 721–726. https://doi.org/10.1109/SYSCON.2015.7116836

Whitney, D., Eldon, M., Oberlin, J., & Tellex, S. (2016). Interpreting multimodal referring expressions in real time. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 3331–3338. https://doi.org/10.1109/ICRA.2016.7487507

CHAPTER 5

Collaborative Object Search

5.1 Introduction

Collaborating with others on shared tasks is a sophisticated and adaptive skill. Collaboration involves multiple individuals working together to achieve a common goal, typically when that goal could not be achieved independently or when it would be very difficult to do so (Hopper & Cronin, 2018). Taken from the start of a task, collaboration may require an individual to evaluate the problem independently, determine if they are unable to solve it on their own, recruit a partner, and then work with that partner in a coordinated fashion to complete the task. The ability to successfully partner in this context requires an ability to form and represent shared goals with another. A collaborator must also maintain some kind of knowledge about their partner's reliability, relationships, as well as their skillfulness (Dale et al., 2020; Hermes et al., 2016).

5.1.1 Shared Intentionality and Human Cooperation

As human beings, we collaborate in play, engaging our social partner in games and sports for social connection and fun. We also work together in a variety of other contexts, teaching, building, and providing care. Some theories of human uniqueness focus on our ability to

collaborate with others, arguing that this sets us apart from other species. One such theory is the shared intentionality hypothesis (Tomasello et al., 2005).

As a highly visual species, our collaborative activities are highlighted by a series of specific visual interactions. The shared intentionality hypothesis emphasizes the importance of visual collaboration between social partners, characterizing three levels of engagement. As human beings develop through these levels, we eventually surpass our closest genetic relatives (chimpanzees) in our collaborative abilities. Critically, the shared intentionality hypothesis defines, at each level, key visual behaviors that identify the level. These visual behaviors are taken as a proxy or at least a behavioral correlate of the developing cognitive abilities of the mind (Tomasello et al., 2005).

We first engage in dyadic engagement. Shared with other species, dyadic engagement involves turn-taking and sharing behaviors. At this level, social partners can share emotions and engage behaviorally without having to represent anything about their partner's mind. In human beings, this ability first develops around 3 months of age. The key visual behavior correlate of dyadic engagement is mutual gaze, which involves simultaneous looking to the face of a social partner (Tomasello et al., 2005).

Human beings begin to leave Chimpanzees behind later in our second year of life. At 9 months of age, we begin to engage in triadic engagement. At this level, partners can share goals, and can integrate their perspective or perceptual cues with those of their social partner. The key visual-behavioral correlate of triadic engagement is visual re-engagement, which involves re-initiation or communication by a social partner when a shared activity is interrupted (Tomasello et al., 2005).

At the start of the second year of life, between 12 and 15 months, children develop the third and final level, collaborative engagement. Collaborative engagement allows partners to understand each other's intentions, plans, and motivations. The key visual behavioral correlate of collaborative engagement is joint attention.

As a cognitive process, joint attention is the process of actively engaging with a social partner while also engaging or interacting with an object. This requires representing the attentional state of the social partner and coordinating your goals to work together in a rich social interaction (Baldwin et al., 2001; Carpenter et al., 1995; Kaplan & Hafner, 2006; Malle, 2008). This suite of cognitive abilities has to be operationalized by researchers, which requires balancing subjectivity and rigor (see review by Bradley et al., 2023). Prior work focusing on the development of joint attention has used questionnaires, and behavioral coding techniques, though these have received criticism for being too subjective and relying too much on the individual researcher attributing intentionality to an observed situation. When efforts are made to be more quantitative using decision trees or visual sequences, the operationalization varies widely across studies (Bradley et al., 2023). For example, one coding scheme requires that a bid for joint attention be intentional, requiring the coder to determine that the initiator act with purpose and "non-randomly". This approach also requires that the initiator verifies that their partner is looking at the item, meaning that they disengage from the item and/or provide a clear acknowledgment of the partner's engagement (Gabouer & Bortfeld, 2021).

Eye-tracking helps to provide quantifiable measures of visual attention, and helps to make definitions of joint attention comparable across studies. Head-mounted eye-tracking in particular allows for naturalistic interactions, capturing the participants' first-person perspective as they move through the world. When tested in eye-tracking studies, joint attention is typically operationalized as the participants looking to an object simultaneously (Abney et al., 2017; Guo & Feng, 2013; Monroy et al., 2020; Yu & Smith, 2013). A major advantage of eye-tracking for studying joint attention is also the capture of subtle shifts of attention. Momentary distractions, or momentary engagement, can both be captured through subtle movements of the eyes that would be very challenging to record from a third-person perspective.

We know that visual behaviors such as mutual gaze and joint attention to objects are critical features of social communication and partnership in human-human interactions (Franchak et al., 2018; Monroy et al., 2020; Yu & Smith, 2013). The operationalization of joint attention is

contentious, as described above. In contrast, mutual gaze is simply looking at one anothers faces (Franchak et al., 2018; Tomasello et al., 2005). There is little debate over how mutual gaze should be operationalized, possibly because this is a behavior that is shared across species and is not theorized to be indicative of rich or complex cognitive abilities (Tomasello et al., 2005).

Joint attention and mutual gaze are often related and can co-occur in close temporal proximity. For example, mutual gaze may end with one person looking at an object. The second participant may notice that the mutual gaze ended and follow the other person's attention to the object. This change in visual attention the second participant makes begins a period of joint attention to the object. Some developmental studies have integrated mutual gaze to operationalize joint attention. For example, Carpenter et al., (1995) counted joint attention as looking at the face and an object within 3 seconds. These stricter definitions are in the minority, and most of the developmental literature uses the simultaneous visual attention described above.

Taken together, by the second year of life, human beings can use joint attention, mutual gaze, and visual re-engagement while participating in collaborative tasks. These visual behavioral signifiers help us to achieve shared goals and intentions. But are they really unique to human beings? Most critically, is joint attention (the visual behavioral correlate of collaborative engagement) truly only seen in our species?

Typical comparisons to chimpanzees are have contributed to a wealth of knowledge understanding where we have diverged from our most recent common ancestor, but they have also been plagued by poor species-normative designs. Chimpanzees have a different ecology than human children, and prior work has sometimes expected atypical behaviors from chimps. This has previously led to underestimations of animal abilities in domains like visual perspective taking. For example, chimpanzees were historically thought to be unable to take the perspective of another. Chimps failed tasks requiring them to selectively beg for food from a human informant with visual access to them vs. an informant who could not see them (Povinelli & Eddy, 1996). Chimpanzees are typically competitive eaters, and begging for food from human informants is not the most ecologically valid test. In a subsequent paradigm where a subordinate chimpanzee

had to compete for food against a dominant chimp, chimpanzees were able to take the perspective of others (seen in this context by the subordinate chimp grabbing the food the dominant chimp could not see vs. the food the in the dominant chimp's line of sight) (Bräuer et al., 2007; Hare et al., 2001). Changing the study design to expect species appropriate behaviors has helped, but to explicitly test theories of human unique cognition we may be better suited by looking at a species that naturally works with us, domestic dogs.

As mentioned in Chapter 1, the natural social partner for the domestic dog is human beings. Dogs and humans work together on various tasks in an applied working context and as pets. In classic tests of collaboration, dogs can successfully work with human partners (Range et al., 2019). Dogs are able to do this flexibly, waiting for their partner through varying time delays. Despite dogs' known success at cooperation, we do not know what features of visual communication promote successful human-dog collaboration and if these features are similar to those seen in human-human teams.

Recent work with dogs has challenged the uniquely human aspect of the second level of the shared intentionality hypothesis, triadic engagement. When there is a break in the activity, humans visually re-engage with each other. Visual re-engagement involves moving into the line of view of the partner or initiating mutual gaze or joint attention to resume a shared activity. Social partners re-engage each other, so when there are breaks in joint attention or mutual gaze, either partner may be the re-initiator. When play is interrupted, dogs re-engage with social partners, a sign of triadic engagement also seen in Bonobos (Byrne et al., 2023; Heesen et al., 2020; Horschler et al., 2022). More specifically, during a play bout with a toy, when the human partner looks away and stops playing, dogs attempt to re-engage the human using a variety of strategies. Like behaviors seen in human children, dogs make eye contact and offer the toy to the partner who was previously playing with them, and they do both of these to the previous player more than they do to a bystander (Byrne et al., 2023; Horschler et al., 2022; Warneken et al., 2006). This is taken as a signal that the previous player and the dog were engaged in a shared goal (play) and is one component of joint intentionality. Re-engagement after the interruption of

activity is, in the view of the shared intentionality hypothesis, stated explicitly to be unique to human beings (Tomasello et al., 2005).

Dogs' ability to selectively re-engage social partners has to develop. Human infants are able to re-engage a social partner from 9 months of age (Ross & Lollis, 1987). By the second year of life, children are robustly able to use re-engagement to facilitate cooperation across contexts (Warneken et al., 2012). Unlike human infants, 8-week-old puppies do not preferentially re-engage a previous playmate (Jonkoski et al., 2026). Described as indiscriminately social, they are equally likely to attempt to engage with a bystander as with a prior social partner when play was interrupted. Further research is still required on the precise developmental timeline of re-engagement behaviors in dogs to understand when this ability emerges. Findings from adult dogs and puppies suggest that what sets human visual re-engagement apart may be the early emergence in development, rather than the behavior itself.

Do dogs engage in joint attention? Given their prior observed success in collaborating with human partners, dogs may display the key visual behavioral correlate of collaborative engagement, joint attention. Moreover, do visual behaviors like mutual gaze and joint attention serve similar functions in this cross-species context? It is possible that dogs display joint attention, as typically operationalized, but that it does not actually indicate any of the cognitive mechanisms of human joint attention. If true, we might see joint attention displayed in visual behaviors but they would not have a relationship to success in a collaborative task. This would provide very strong evidence for the need to update how we operationalize joint attention, but would still support the shared intentionality hypothesis broadly. Alternatively, we could see joint attention in human-dog interactions, and it could help to facilitate cooperation. If observed, this would provide evidence for a rewriting of the shared intentionality hypothesis. Specifically, the ability to engage in the key visual behavioral correlate of collaborative engagement would no longer be unique to our species. Remedying this could involve a stricter operationalization of the visual behavior of joint attention, or the creation of a scale or gradient of joint attention with human-dog joint attention at one end and full adult human joint attention at the other. Either

way, the findings of this study will have significant implications for our theories of what sets us apart from other species.

5.1.2 Factors Influencing Collaboration

People cooperate and collaborate with each other on the basis of a variety of factors. When deciding whether to work with someone, and the extent to which we want to collaborate with that person, we may consider our emotional connection to our candidate partner (Huang & Lajoie, 2023). In human collaboration, the social-emotional bond plays a vital role. We prefer to work with those we feel more bonded to, but we do not only extend emotional bonds to human beings. People bond to varying degrees with their dogs, perceiving different levels of relationship closeness. To examine if this had a potential impact in this cross-species context, we gathered survey data from dog owners on their perceived emotional bond with their dogs (Dwyer et al., 2006; Howell et al., 2017). We then examined if dog-owner perceived bond was correlated to task success.

Another major factor in efficiency of and willingness to collaborate is perceived intelligence. The perceived intelligence of a social partner or agent is a frequent topic in the robotics literature and in human-machine interaction studies (Gao et al., 2025; Liao et al., 2024). If we perceive a candidate partner as intelligent we are more likely to want to work with them, seeing the collaboration as more advantageous. Perceptions of intelligence extend not only to human beings but also to robots. People make judgments about perceived dog intelligence, but does this impact collaborative success? To test this we gathered data on dog owners' perceptions of their dogs' intelligence prior to task participation to determine if there was a relationship between perceived intelligence and task success (Pelgrim et al., 2026; Ross et al., 2024).

In this study, we aimed to explore, using dual-eye-tracking on an ecologically valid task, the visual behaviors in human-dog teams (hereafter dyads). Researchers have often used dual eye-tracking paradigms to study the visual behaviors behind human interactions. In these setups, participants (typically parents and children) wear head-mounted eye-trackers while interacting

with stimuli. These paradigms often emulate natural object-oriented play. Work using these systems has uncovered how children and parents orient their attention in an ecologically valid context (Monroy et al., 2020; Yu & Smith, 2013). Dual eye-tracking paradigms have also revealed the importance of joint attention in a pedagogical context (Shvarts, 2018) and how adults coordinate their attention in collaborative games (Jermann et al., 2010).

Joint attention and mutual gaze predict success in cooperative tasks in human teams. We examined if key visual behaviors served a similar role (meaning that they promote cooperation) in this cross-species context. We expected to see both visual behaviors in dyads and hypothesized that these visual interactions served a similar purpose in dyads as in a human-human context. We anticipated that increased instances of mutual gaze and joint attention would predict success on the object search task. We also explored if the dog-owner perceived intelligence and dog owner perceived emotional bond related to success on this cooperative task.

5.2 Methods

5.2.1 Participants

The participants were 13 dog-owner pairs. 13 dogs (6 = Female, 7 = Male, Mean Age = 64.70 Months, Age Range = 14.59 - 127.33 Months) and 13 adults. Our human participants self-reported gender as 10 “Female” and 3 “Male”. They self-reported race/ethnicity as 9 “White”, 2 “Asian”, 1 “Black or African American and Asian” and 1 “Latino or Hispanic and White”. Dogs were recruited from the eye-tracking goggle training program at Brown University. An additional 4 dyads were excluded due to (1) Equipment failure resulting in data loss (N = 2), and (2) failure of the dog to wear the goggles comfortably (N = 2)

5.2.2 Ethics Note

All ethics guidelines were followed, and study procedures were approved by the relevant ethics committees. All components of the study involving dogs were approved by the Brown

Institutional Animal Care and Use Committee, Protocol Number 23-09-0002. Components of the study involving adult humans were approved by the Brown Institutional Review Board Protocol Number STUDY00000124.

5.2.3 Materials

Eye-Trackers

During our experiment, dogs and their guardians wore wearable eye-trackers for the duration of the session (Figure 5.1). Dogs wore a custom-developed head-mounted eye-tracker, where two cameras were affixed to commercially available dog goggles (Positive Science, Inc.). This system has previously been used and validated in dogs (Pelgrim et al., 2022, 2025). In line with past work, in addition to the eye-tracking goggles, dogs wore a harness throughout their session to hold the recording and battery packs. Dog guardians wore eye-tracking glasses (Positive Science, Inc.) and carried their recording and battery packs in a clear cross-body bag. The eye-trackers also record audio for the duration of the session. Vocalizations from either the dog or the guardian were picked up from their worn eye-trackers.

Room Set Up

The room setup (Figure 5.2) was identical for all dog-guardian pairs. The room contained objects of different sizes and were located at different heights to serve as potential hiding locations. Locations were classified either as dog-accessible, meaning that dogs are physically able to access and consume a treat hidden there independently, or dog-inaccessible, meaning that dogs are physically unable to access the treat without human intervention. Dyads were assigned to one of two hiding arrangements that determined the precise locations of the hidden treats. The order of hiding was randomized for each dog so that test trial accessibility types alternated.

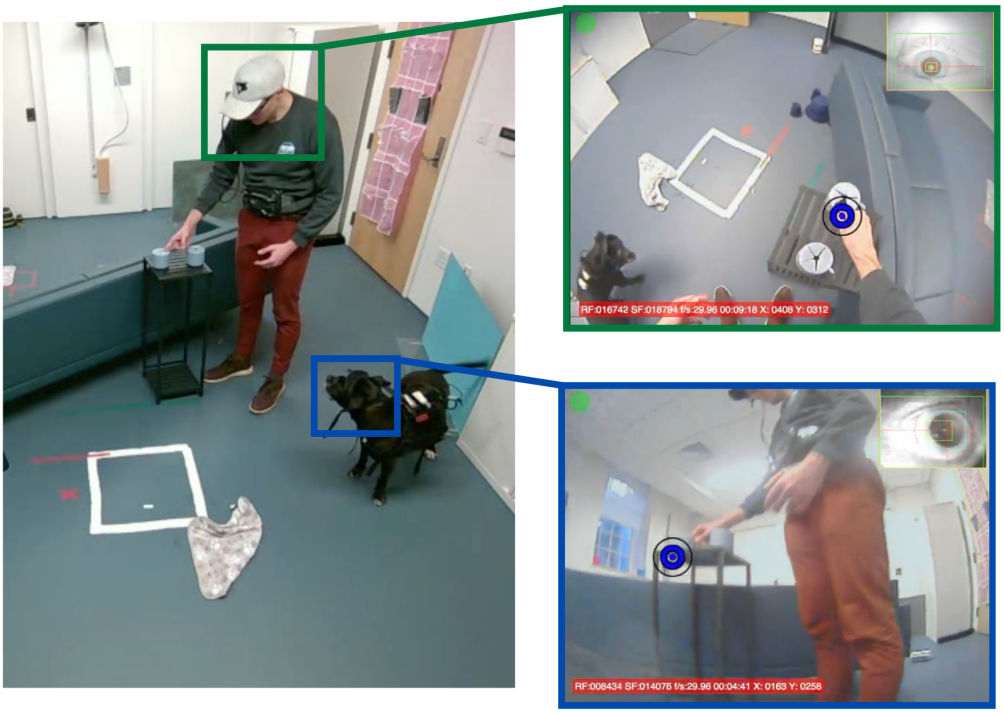

Figure 5.1: An example of participants searching a cup at the end of a trial. Both participants (Left) are wearing eye-trackers and we synchronize the human's perspective (Top Right) with the dogs' (Bottom Right). This is also an example of joint attention, with both participants simultaneously looking at the same object.

Guardian Surveys

Prior to attending their session, dog guardians completed the Monash Dog Owner Relationship Scale (MDORS) (Dwyer et al., 2006; Howell et al., 2017) and the Dog Intelligence Perceptions Survey (Pelgrim et al., 2026; Ross et al., 2024). If dog owners had previously completed the Dog Intelligence Perceptions Survey as part of a prior unrelated study, they did not complete the survey a second time. Their original responses were used.

5.2.4 Procedure

At the start of the session, participants were fitted with their eye-tracking equipment and they completed a calibration procedure. During the procedure, a participant followed an item of interest with their eyes. Dogs were calibrated using a treat, and their guardian held their head

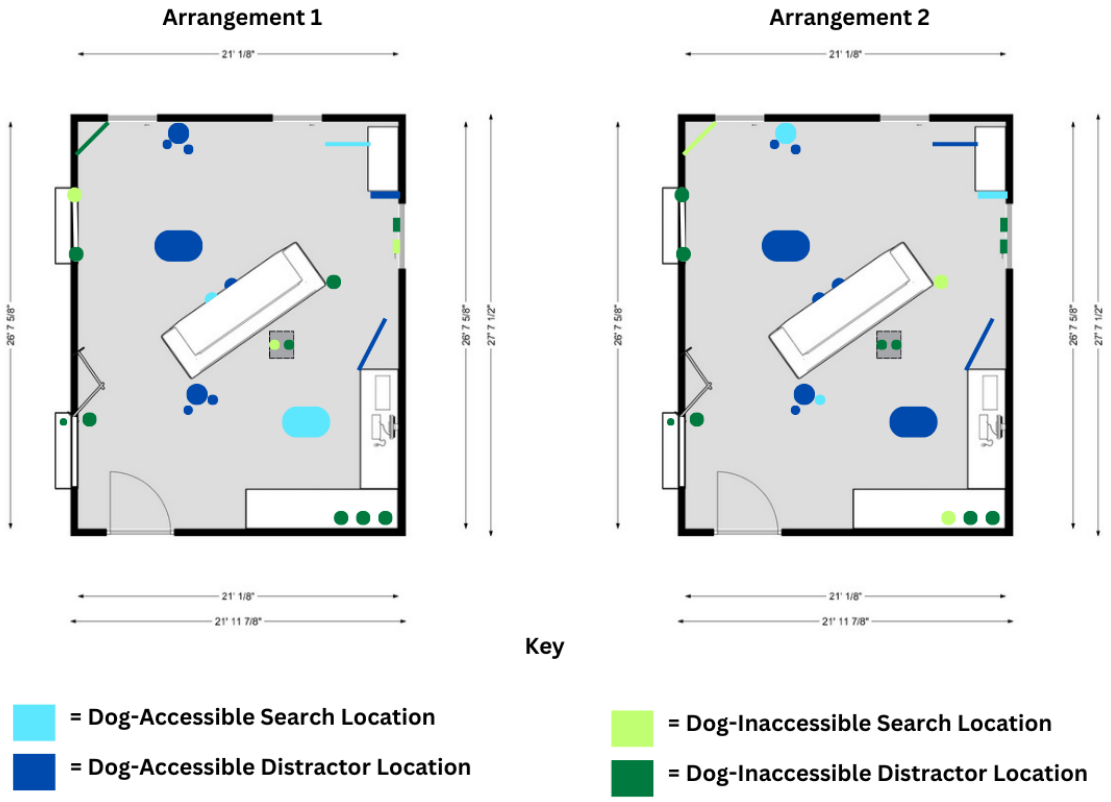

Figure 5.2: The items in the room were arranged identically for all participants. Participants were randomly assigned to be in one of two hiding arrangements.

approximately still. Guardians were instructed to not move their head and watch a small toy. If at any point during the session there were concerns about a disruption of the eye-tracker, a portable LCD screen was used to verify a good eye-image. If the eye camera had been disturbed, a calibration procedure was repeated. In the event of other disturbances (i.e., hair fell into the eyes disrupting the eye-image) the situation was resolved without the need for additional calibration (i.e., hair was removed from the face and re-tucked within the goggle frame).

During test trials, the experimenter hid a treat while both participants were in the waiting room. Depending on the trial accessibility type, either the dog or the owner watched the hiding and the other member of the pair did not observe. In dog-led search trials (3 total), dogs were able to view the hiding of the treat in a dog-inaccessible search location while their guardian could not see. To find the treat, dogs had to recruit their owner to get the treat from the inaccessible

location. In guardian-led search trials (3 total), guardians viewed the hiding of the treat in a dog-accessible search location. To help their dog find the treat, guardians had to engage their dog and direct them to the accessible location to search there. At the start of all test trials, dogs were given a 5-second head start to begin the trial while guardians could watch and speak but not enter the room. This allowed dogs to begin to attempt the task independently, though this behavior was rare and the majority of dogs waited in the waiting room for their owner. After 5 seconds, guardians were permitted to enter the rooms. Dyads had a total of 2 minutes to search for the hidden treat.

After test trials, dogs completed two control trials. During these, dogs were able to view the hiding, and the treat was hidden in a dog-accessible location. These trials could be solved entirely independently by the dog and did not require but could contain cooperation or engagement with their guardian. This helped us to assess baseline levels of joint attention or mutual gaze in the absence of information seeking. If these visual behaviors were present at the same rate in the control trials as in our test trials, it would suggest that the visual behaviors are not serving a similar role to that seen in humans. As with test trials, dogs were given a 5-second head start to begin approaching the task. The session concluded with a final calibration procedure for both dog and guardian.

5.3 Data Coding

5.3.1 Behavioral Data

During the session, dyads were recorded as successful or unsuccessful at finding a treat on a given trial. Using video recording after the session, the timing for successful trials was determined.

5.3.2 Eye-Tracking Data

After the session, videos from the head-mounted eye-trackers were rendered into usable eye-tracking data using Yarbus eye-tracking software (Positive Science, Inc.). This identified the timing (start and stop times) and direction (x-y coordinates) of participants' fixations. In keeping with prior work, fixations were defined as stable eye positions lasting at least 100 ms (Pelgrim et al., 2022, 2025). For each fixation that occurred during trials, the item the participant fixated on, hereafter referred to as the target of fixation, was manually coded using Gazetag software (Positive Science, Inc.).

The accuracy of the rendered footage was assessed for each participant. In keeping with past work, we used supplemental points from our calibration procedure that had not been used to generate the rendered files (Pelgrim et al., 2022, 2025). This allowed us to determine the difference between the estimated fixation target (as generated from the rendered eye-tracking) and the true fixation target (the item being held by the experimenter). We used a total of 20 frames across the visual field of the participant to determine average calibration accuracy.

Fixation targets were defined as one of 17 possible search locations (used for all joint-attention analyses), the partner (used for all mutual-gaze analyses), the experimenter, the floor, or off-target. Each fixation was coded by two independent coders (MHP and one of two research assistants). In the event of disagreement between the coders, coders met and came to a consensus.

We synchronized eye-tracking data across participants manually. After synchronization, we examined fixation target data for specific visual behaviors of interest. First, we identified instances where dogs and their guardians looked to each other, termed mutual gaze. We operationally defined mutual gaze as instances where the dog and the guardian were simultaneously fixated to the ROI "partner".

Our second key visual behavior was joint attention. In line with past work (Yu & Smith, 2013), we operationalized joint attention in two ways, synchronized and sustained joint attention. Synchronized joint attention we defined as any individual fixation where both partners were

fixated on the same ROI. Sustained joint attention captured instances of repeated synchronized joint attention. In keeping with prior work from children (Yu & Smith, 2013) we defined sustained joint attention as instances of synchronized joint attention that were separated in time by less than 300 ms and in total summed to over 500 ms. We acknowledge that

When there was a look at the same location/item by both individuals (synchronized joint attention), we determined who (dog or human) looked at the item first and how long it took for the partner to look at that item. This allows us to determine, for each instance of joint attention, who the leader of the visual interaction was and how much they led. We called the time between the look to the item by the leader and by the follower the lag time.

As an exploratory measure, we also looked for two additional visual behaviors, referential gaze and gaze alternation. We operationalized referential gaze as looking to the partner within 2 seconds of a fixation to a potential search location. We also captured instances of gaze alternation, operationalized as a three-part series of fixations involving a look to the partner, an object, and the partner again (or object, partner, object again), all occurring within 2 seconds. These kinds of visual referential communications have previously been observed in dogs through recordings of head movements and estimations of gazing behaviors, but have not been seen in eye-tracking data (Marshall-Pescini et al., 2013; Merola et al., 2012; Pelgrim et al., 2024).

5.4 Results

5.4.1 Survey Data

On a scale of 0-100, dog owners in our sample viewed their dogs as above average in intelligence, M = 78.15, SD = 14.87, Range = 33 - 90. Using the complete survey, dog perceived intelligence can be broken down along the three sub-components, skill in social reasoning, challenges with physical reasoning, and temperament (approach to novel situations and general executive function skills). We averaged across the items to generate a score per component for each participants, and scores had a potential range of 0-100. Owners reported their dogs as

having few problems with physical reasoning (M = 25.5, SD = 11.9, Range = 12.2 - 53.3). In contrast, owners in our sample reported their dogs as being skilled at social reasoning (M = 62.3, SD = 12.2, Range = 41.6 - 80.4). Finally, owners in our sample reported more modest ratings on our temperament component (M = 51.7, SD = 9.84, Range = 37 - 74).

Dog owners in our sample were highly bonded to their dogs, and perceived a close relationship as measured by the MDORS (Dwyer et al., 2006; Howell et al., 2017). The MDORS contains three sub-scales and participants are given a score for each sub-scale by averaging the individual items within it. Some items are reverse scored so that the scale can consistently be read on a 1-5 with higher scores indicating a higher perceived relationship quality. From our sample, dog owners perceived the costs of owning their dog as low (M = 4.12, SD = 0.79). Owners also reported high levels of interaction with their dogs (M = 3.98, SD = 0.55) and perceived high levels of emotional closeness (M = 4.06, SD = 0.65).

We explored if dog-owner perceived bond (as measured by MDORS) and dog-owner perceived intelligence correlated with performance on the object search task. It was possible that dog-owner teams with closer owner-perceived bonds would be more successful. However, because of self-selection in our sample, we have little variance in owner-perceived bond and intelligence. With the exception of one participant (who rated dog intelligence as a 33 on our scale of 0-100), participants generally believed their dogs to be highly intelligent, have few problems with physical reasoning, be highly skilled at social reasoning, and have a moderate temperament. We found no statistically significant correlation ($p > .05$) between any of our measures of dog intelligence and average dyas task performance. Dog owners also perceived the costs of owning their dogs as low, and interacted and felt emotionally close to their dogs. Similarly, we found no statistically significant ($p > .05$) relationship between the total MDORS score or any of the subscales. Conclusions from these findings should be limited due to the minimal variance in our survey data. It is possible that bond and perceived intelligence to play a role in human-dog collaboration, but that we were not able to detect the effects here.

5.4.2 Behavioral Data

Dyads successfully found the treat on 70/78 test trials (89.74%). When they were guided by their guardian, all dogs were successful, finding the treat on 39/39 test trials (100 %). While still successful, dyads were less likely to find the treat when dogs knew the location. During dog-led search trials dyads successfully found the treat on 31/39 trials (79.49%). Similarly, the average time to completion in seconds was lower for owner-led searches (M = 17.4, SD = 6.70) than for dog-led searches (M = 42.8, SD = 31.8). All participants found the treat on all control trials, 26/26 control trials (100 %). Control trials were also solved quickly (M = 8.59, SD = 7.03). Half of the control trials (13/26) were solved by dogs during their head-start phase (the first 5 seconds) prior to the owner entering the room.

5.4.3 Eye-Tracking Data

Calibration accuracy was comparable to that seen in prior work. Human participants had an average calibration accuracy of 2.82°. Dog participants had an average calibration accuracy of 3.18°. Across the sessions, coders initially agreed on 86.07% of fixations. The coders resolved all instances of disagreement for the codings present in this analysis. Across our trials we had a total of 9,352 fixations used for analysis.

Our two key visual behaviors of interest were mutual gaze (simultaneous looks to each other) and joint attention (simultaneous looks to an object). Individual bouts of mutual gaze tended to be short (M = 195.45 ms, SD = 0.29) and were observed in 13/13 pairs and on 53/78 individual test trials. On trials where mutual gaze was present it, it took up 8% of total fixation time. In comparison to prior work with human children, individual bouts of mutual gaze last for a shorter period of time between people and dogs; however, they are occurring for a comparable amount of total trial time (Yu & Smith, 2013). Bouts of synchronized joint attention tended to also last be short (M = 178.67 ms, SD = 0.17). Joint attention bouts were observed in 13/13 pairs and on 54/78 individual test trials. On trials where joint attention was present, participants spent an average of 9% of the trial jointly attending to a search location.

We conducted a mixed-effects regression to explore if our two key visual behaviors (mutual gaze and joint attention) predicted task success. Specifically, we explored how the proportion of fixations from a trial engaged in mutual gaze, the proportion spent in synchronized joint attention, and a potential interaction between the two predicted the logged time to success on each trial[1]. For the purpose of this analysis, we removed the 8 unsuccessful trials as those trials did not have a time outcome. We also included a random effect per dyad.

The proportion of mutual gaze engaged in during a trial had a significant main effect on time to task success, $F(1,519.55) = 14.23, p < .001$. Dyads who engaged in more mutual gaze tended to have shorter trial times, meaning they were successful faster. Likewise, the proportion of time engaged in joint attention had a shortening effect on logged trial time, $F(1,507.93) = 28.94, p < .001$. Finally, there was a significant interaction between the proportion of time spent in joint attention and mutual gaze in predicting time to trial success, $F(1,517.91) = 38.616, p < .001$. We found that the amount of mutual gaze could speed up or slow down dyads depending on the amount of joint attention on that trial. Dyads who engaged in longer amounts of joint attention and longer amounts of mutual gaze had slower overall trials. In contrast, dyads that engaged in shorter joint attention and longer mutual gaze had shorter overall trials (Figure 5.3). This suggests that the relation between collaborative or shared looking patterns and task success may be an inverse U-shaped curve. Some level of collaborative looks, be it mutual gaze or joint attention, help dyads to coordinate and locate the hidden item. However, too much collaborative looking slows them down and may hinder more than it helps. We also saw a moderate level of variance explained by dyads from the random effect in our model ($ICC = 0.376$).

The observed relation between joint attention and mutual gaze is also consistent across trial types. When considering dog-led trials and owner-led trials separately, we see the same pattern of results presented above. This suggests that the interactive pattern of joint attention and mutual gaze is one that is indicative of dog-human interactions generally. If we had only

[1]We used logged trial time for this analysis because raw trial time violated heteroscedasticity. We also explored dropping the outliers that were driving this violation (all outlying trials were from a single dyad), incorporating clustering, and dropping the outlying cluster. All of these resulted in the same pattern of results presented below.

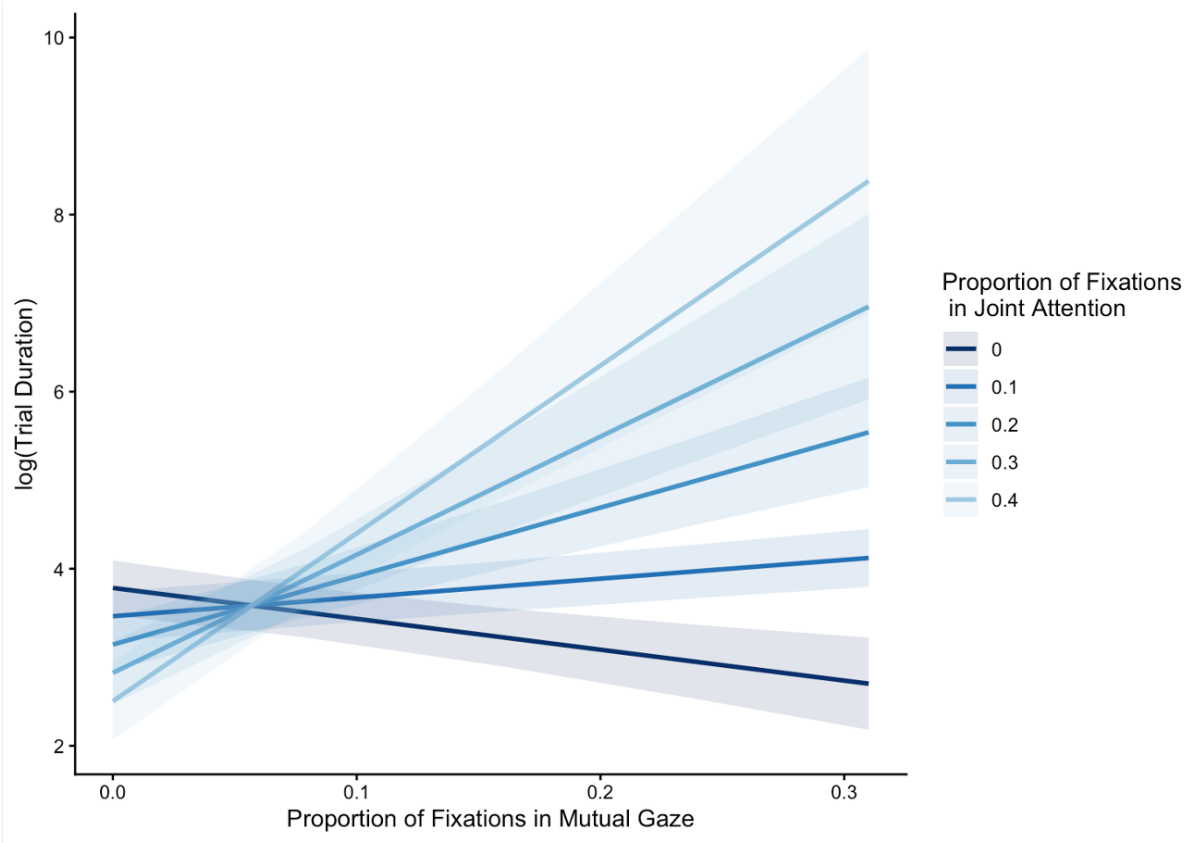

Figure 5.3: Relationship between the proportion of trial fixations engaged in mutual gaze and joint attention on logged time to success.

seen a relationship between joint-attention and task success (for example) on owner-led trials, it would be likely that the patterns of visual behavior were being driven strictly by the owner who knew the information and had all of the necessary context.

In keeping with prior eye-tracking work (Pelgrim et al., 2022; Yu & Smith, 2013), levels of our two key visual behaviors could also be operationalized as proportion of total trial time rather than proportion of trial fixation time. This is a subtle difference, but proportions of trial time are typically smaller than proportions of fixation time. By the definition of a fixation (a stable look lasting at least 100 ms) there are instances of transition where the eye moves from one fixation to the next, blinks, and other brief sequences of time without any fixation data. By calculating both ways, we can feel confident that the pattern being reported is robust and not being over-inflated due to tracking losses for reasons cited above. When we calculated joint attention and mutual gaze as a proportion of total trial time and ran the above model, we found

an identical pattern of results to those reported above. This further confirms that mutual gaze and joint attention are impacting dyad time to trial success.

We can also measure sustained joint attention (instances of repeated synchronized joint attention lasting $>$ 500 ms). Sustained joint attention towards objects in the room (M = 885.03 ms, SD = 0.43) was observed in 12/13 pairs and in 23/78 individual test trials. When sustained joint attention occurred, it took up an average of 13% of the total fixation time for that trial.

We have now demonstrated that joint attention and mutual gaze are having a significant impact on task success. Further, we observed joint attention in both dog and human led trials, being initiated by both members of the dyad. What if this visual coordination, or these instances of joint attention, are occurring by chance? In keeping with past work (Yu & Smith, 2013), we examined the lag times from our sample. We anticipated that dogs and humans were visually coordinating with each other, so we expected that the distribution of lag times from our sample would be significantly different from a random baseline.

To test this, we temporally re-ordered fixations separately for each member of the dog-human dyad so that the order of the fixations was random. We then compared if the randomly shuffled cross-lag was significantly different from the original temporally accurate data using a mixed effects model. To maintain comparability to past work (Yu & Smith, 2013), we binned the lag times from both the real and randomly shuffled data into 61 bins[2]. We then ran a mixed effects linear regression predicting the density of the bins as a function of the data type (real vs. randomly shuffled), the identity of the bin (1-61 with 1 indicating maximum lead time by the human and 61 indicating maximum lead time by the dog), and an interaction between the two. If there was no "true" visual coordination and our identified instances of joint attention were happening randomly, then we would not expect to see a statistically significant difference between the real and randomly shuffled data. Alternatively, if dyads were in engaging in joint attention and coordinating their gaze in real time, we would expect to see significantly higher

[2]61 bins was chosen because it was used in prior work. Changing the number of bins (tested ranging from 5 to 125 bins) does not significantly change the pattern of results

levels of joint attention (measured as bin density) in our real data than in the randomly shuffled data.

As seen in Figure 5.4, there were many more instances of joint attention in our real data than in the randomly shuffled data. We found that the density of the bins in our real condition was significantly higher than those in the randomly shuffled version, $F(1, 1452) = 1156.96, p < .001$. This provides strong evidence that dogs and humans were coordinating their visual attention while completing our task. The instances of joint attention observed in our sample are not occurring randomly, confirming our observation of joint attention in a non-human animal.

Similar to past work, we saw a sharp peak near the center, indicating that lag times tended to be quite small or that gaze following into joint attention was relatively quick (Figure 5.4). Unlike in past work our peak was slightly offset (Yu & Smith, 2013). From the same model, we also found a significant effect of the lag bin, $F(60, 1452) = 31.028, p < .001$. The highest density averaged across our data was for bin 32. This suggests that the distribution of our lag times was skewed towards dogs leading, but that both species did initiate or lead joint attention.

In addition to occurring far more frequently, our real observations of joint attention had different lag time frequencies than the randomly shuffled data. From the same model we also saw an interaction between lag bin and condition, suggesting that the distribution of the cross-recurrence lag differed between the real and shuffled, $F(60, 1452) = 15.62, p < .001$. See Figure 5.4 for an unbinned view of the distribution of lag times. This provides further confidence that the instances of joint attention we see in our real data are not only happening more often than random chance, but they are also happening with a different structure than random chance. We saw negligible variance explained by participant from the random effect in our model (ICC = .029).

We also examined differences in recurrence from a shuffled or randomized baseline using a cross-recurrence quantification analysis (CRQA) (Coco & Dale, 2014; Wallot & Leonardi, 2018). This allows us to go beyond a frequency or density approach, and examine broader connections

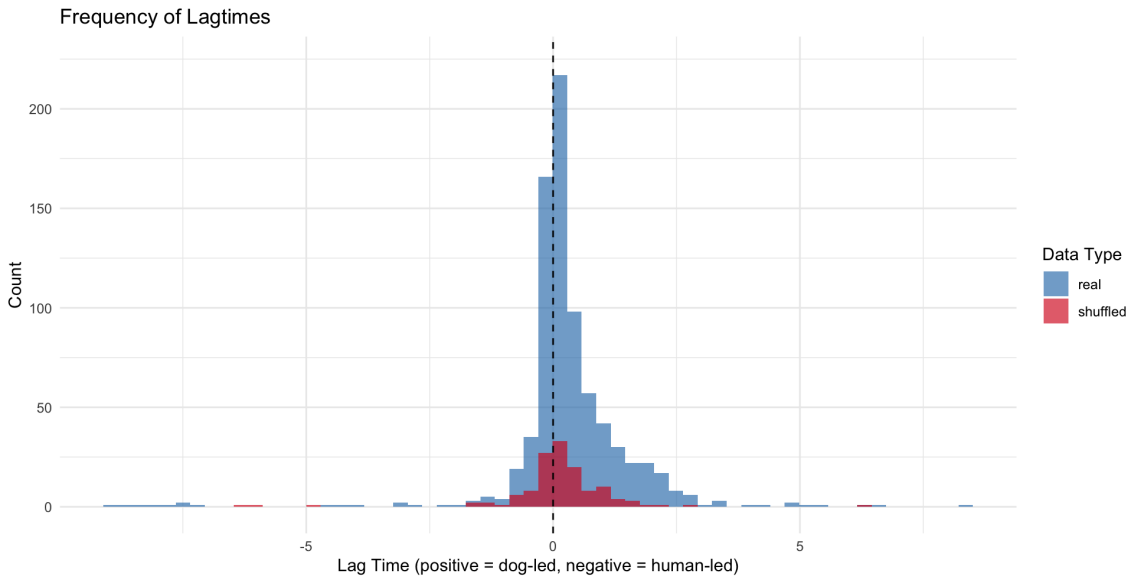

Figure 5.4: Lag time frequencies for instances of joint attention for real data and for randomly shuffled data. If shuffled and real data were the same, the two areas would be of equal size. If dyads were leading equally data would be symmetrical around the black hashed line.

across fixations. We calculated the percent determinism ($\%DET$), or how many of the instances of joint fixation targets occurred in connected trajectories. As with our binned analysis, we created a random baseline by shuffling the order of our fixations within our participants, and then compared $\%DET$ across the real and shuffled calculations. We ran a mixed-effects linear regression comparing the average outcome per participant as a function of data type (real or randomly shuffled) and a random intercept per dyad. We found a significant difference in $\%DET$ between the real and randomly shuffled conditions, $F(1,12) = 961.47, p < .001$. This provided further confidence that dyads were coordinating their visual attention while solving our task. Finally, from our same model, we saw little variance explained by dyad from our random effect ($ICC = 0.109$).

Dyads displayed instances of joint attention and mutual gaze. To better understand this pattern qualitatively, we reviewed if visual leading by dog or human predominates, or whether they're symmetrical like in parent-child interactions (Yu & Smith, 2013). We examined this both by the occurrences (counting the instances of dog or owner led joint attention bouts) and by the onset time (measuring how long the leader was looking to the item before the follower).

We found that dogs generally led instances of joint attention. Of the instances of joint attention in our sample, dogs led the looks 399/559 (71.37%) of instances. Lag times in seconds were generally larger when dogs led joint attention (M = 0.65, SD = 0.79), than when owners led (M = 0.21, SD = 0.28). Dogs also tended to lead instances of mutual gaze, leading on 229/382 (59.95 %) of instances. Unlike bouts of joint attention, dyads tended to lead by equal amounts of time when engaging in mutual gaze. Dogs (M = 0.81, SD = 1.09) and owners (M = 0.87, SD = 1.85) both had larger lag times for instances of mutual gaze than they did for joint attention. Across our test trials, we generally did not see symmetrical lag times. This is an important difference from typical parent-child interactions. It suggests that although joint-attention may serve a similar role in search tasks in both species, the initiation of these behaviors is different in our cross-species context.

On half of the trials, dogs had the information about the location of the treat, and on half of the trials their owners had the information. We predicted that the individual who knew the location of the hidden item would generally lead the gaze. We conducted a mixed-effects linear regression predicting lag time as a function of the trial accessibility type (or whether the dog or human had knowledge about the location of the hidden item). We found that when guardians had knowledge about the location of the hidden treat, we saw much larger positive lag times (M = 0.903, SD = 1.19) than when dogs knew the location (M = 0.163, SD = 1.32). From our regression we saw a significant main effect of trial accessibility type $F(1, 771.13) = 17.04, p < .001$. This suggests that even when the dog owner knew the location of the treat, they still tended to follow their dogs' visual lead. This finding is interesting but we cannot draw definitive conclusions here. It is possible that owners are "following" their dogs' visual lead during these trials to check if their dog is paying attention to the verbal or gestural cue they are providing. It is also possible that the task is more enjoyable or interesting for the owner if they allow their dog to explore the area ahead of providing instructions. Alternatively, owners may not believe their dog will follow their visual attention so they take on the role of visual follower even when they have the information. Ongoing work, mentioned in the discussion, can help to parse apart these potential

interpretations. We also saw little variance explained by dyad from the random effect in our model ($ICC = 0.21$).

We predicted that during our control trials (when the task can be fully solved independently), dyads would engage in fewer instances of mutual gaze and joint attention than when the task must be solved collaboratively. This prediction was well supported by our data. We saw one instance of mutual gaze across all 26 of our control trials. Only 3 of our dyads engaged in joint attention during our control trials. From these three dyads, we saw 10 total instances of joint attention on our control trials. This rate is much lower than what we saw during our test trials. This suggests that dyads collaborate adaptively. Joint attention and mutual gaze facilitate cooperation and collaboration, but they are not without effort. When dogs are capable of solving tasks independently, they are less likely to try to engage their owners in these collaborative visual behaviors. Similarly, when dog owners know their dog can solve the task independently they do not try to initiate collaborative visual behaviors.

As an exploratory measure, we also examined our data for two other visual behaviors, referential gaze and gaze alternation. We defined referential gaze as looks to the partner within 2 seconds of a fixation to a potential search location. We defined gaze alternation as a three-part series of fixations involving a look to the partner and a search location. Gaze alternation could either follow a partner first approach (partner-item-partner) or a search-location first approach (item-partner-item). All of these looks had to occur within 2 seconds. These two visual referential communication behaviors have previously been observed in dogs and other species through recordings of head movements and estimations of gazing behaviors (Gaunet & Deputte, 2011; Merola et al., 2012; Smith & Litchfield, 2013). Therefore, we expected to see both referential gaze and gaze alternation, but this was the first effort to demonstrate them using a wearable eye-tracking approach. From our sample, we saw all 13 of our dyads and all 26 of our participants engage in referential gaze during our test trials. We saw a total of 614 instances of referential gaze. The majority of our instances of referential gaze (417/614 or 67.91%) were done by dog owners.

Similarly, each of our dyads and all but one participant engaged in gaze alternation. Gaze alternation was also predominantly led by owners (161/224 or 71.88%). Most of our instances of gaze alternation (141/224) were partner-first, meaning that the participant first looked to their partner, then looked to a search location, then looked back to their partner.

In keeping with our pre-registration, we have used operationalizations of joint attention from prior work using a dual eye-tracking method with parents and children. However, there are multiple operationalizations of joint attention in the literature with some increasing in complexity and emphasizing specific visual sequences. Joint attention can also be operationalized as a sequence combining looks to an item in the environment and looks to the partner. For example, joint attentional episodes can be operationalized as a look at the social partner, accompanied by a look at the item of interest in 3 seconds prior to or after the look at the partner (Carpenter et al., 1995). Episodes ended when the individual had not looked at the item or partner for over 3 seconds. Under this approach, we saw joint attentional episodes in 13/13 of our dyads and 77/78 individual trials.

Joint attention could also be operationalized with a stricter sequence-based (vs. time-based) approach (Malle, 2008). Specifically, the leader first looks at an item. Next, the leader looks at their partner to see if their partner has followed their gaze to the target. If the partner is looking at the item, the leader returns their gaze to the target, and this is considered joint attention. If the partner is looking at something else, the leader will repeat the process. In our sample, we saw instances of these joint attention sequences in 13/13 of our dyads and on 32/78 individual trials.

5.5 Discussion

In this chapter, we examined how people work together with their dogs to find hidden items in a complex environment. Dog owners in our sample were highly bonded to their dogs and viewed their dogs as intelligent. We found that dyads are generally successful, recruiting their partner to retrieve the hidden treat. However, people were more successful at recruiting their

dogs (succeeding every time) than when dogs had to recruit their people (succeeding 31/39 times).

We found that all dyads engaged in joint attention, something that should have been uniquely human (Tomasello et al., 2005). Further, we found that joint attention and mutual gaze levels play a role in dyad success, with increased levels of each generally predicting a shorter time to success. Mutual gaze and joint attention levels interact, suggesting that too much collaborative or shared looking results in worsened performance. Dyads also engaged in sustained joint attention. Outside of eye-tracking contexts, joint attention has been operationalized through behavioral gaze coding and through questionnaires. A major challenge of these is the requirement that bids for joint attention be "intentional" (Bradley et al., 2023). Under the shared intentionality hypothesis, dogs do not form joint intentions with human beings. As such, it is impossible for their bids for joint attention to be categorized as intentional. The requirement of these approaches for the coder to make a subjective judgment about the motivations and intentions of the participants make testing joint attention in other species nearly unfalsifiable. If we allow that dogs can have intentions and potentially share these with others, we do not have to say that they are doing it in the same way as human beings. We can, though, then test out using further operationalizations of joint attention with our existing sample to see how many of them would also be demonstrated in our dyads.

Dyads coordinated their visual attention to targets, but unlike prior work with human-human pairs (Yu & Smith, 2013) we did not see symmetrical leading. Specifically, we saw that dogs were typically leading, or initiating, instances of joint attention. This may be because dog owners were more likely to follow the gaze of their dogs because of their dogs' relatively limited communicative repertoire. Relative to human children, dogs have fewer communicative avenues to coordinate with their partners. Dogs cannot use gestures like pointing, nor can they use language. Body orientation and gaze are the primary ways dogs can communicate their attention. Dog owners know this about their dogs, and because of it may be particularly attentive to follow their dog's gaze. This further indicates that the cognitive processes underpinning joint attention

are likely different in human-dog dyads (as compared to human-human dyads). Regardless though, the specific visual behavioral correlate has now been seen in this species. Further information about the cognitive processes underlying our observed joint attention behaviors can be gained from looking at other visual and auditory behaviors from our sample.

In contrast to joint attention, instances of gaze alternation and referential gaze were dominated by dog owners. This may be indicative of attention checking or confirmatory looks. Although both species engaged in these visual behaviors, they were much more common for the human member of the dyad. This could suggest that much of the "higher-level" cognitive processes are being driven by the human, and that the joint attention we are seeing, using published operationalizations, is more indicative of gaze following than true cognitive joint attention.

Joint attention may be a spectrum of cognitive abilities rather than a binary one. Dogs' and children's cognition in this regard may not diverge until later in development. Further research is needed to better identify precise timelines, but it is possible that children and dogs are relatively similar in their joint attentional abilities at 12 months, but by 15 months, children have continued to develop, and dogs have not. Alternatively, dogs may have an entirely different cognitive mechanisms that is facilitating visual joint attention and collaboration. If true, dogs might follow a different developmental trajectory for the observed joint attention presented here than the three levels or stages predicted by the shared intentionality hypothesis (Tomasello et al., 2005).

5.5.1 Future Directions

The results presented here provide compelling evidence that joint attention, as we have classically operationalized it, is not unique to our species. It serves a similar function in collaboration, facilitating task success. This finding opens up many avenues of future research, including ongoing work that is beyond the scope of this dissertation.

Future work might consider further examining the behaviors displayed by dogs and humans during the task. From pilot data, there may be evidence of primacy/recency effects particularly in dogs' search behaviors. More specifically, dogs often first searched the last location they were

rewarded at, even in instances where they observed the hiding of the treat. Additionally, we can further explore the timing of physical searching behaviors as they relate to our key visual behaviors. When joint attention or mutual gaze occurs relative to the overall trial time could have different practical effects.

As part of our dual eye-tracking, we recorded audio from our participants. Unlike what we observed in our pilot, our test sample contained no vocalizations from dogs. In contrast, we heard extensive vocalizations from dog owners directed to their dogs. Future work can consider coding both the timings and content of the speech and vocalizations produced by owners. Future work might explore if vocalizations are synchronized and/or related to the visual behaviors. For example, owners may tend to vocalize the dog's name prior to mutual gaze bouts. Incorporation of vocalizations have historically been used in some measures of joint attention in human children. An intentional bid for joint attention could begin with a vocalization, and the checking behaviors required by certain operationalizations can also be accomplished with vocalizations (Gabouer & Bortfeld, 2021). Exploring this multi-modal communication can provide additional insight into the collaborative process in dyads.

We also recorded our participants using depth tracking cameras. In ongoing work, we are examining the body positioning and orientation of our participants throughout test trials. This can potentially be used to estimate the field of view of the participants, and be synced with auditory codings. Owners may, for instance, be more likely to vocalize when they are oriented towards their dog, but their dog is oriented away from them. In this ongoing work we are also examining how the proximity of the participants changes across trials. We can integrate this with the eye-tracking data to examine where joint attention behaviors/mutual gaze occur relative to space in the room and relative to space between each other. Finally, we are using skeleton tracking of our human participants from our depth data to examine how and where pointing gestures may be used, and will relate these findings to our eye-tracking data and proximity data. Dogs are physically unable to produce many prototypical human gestures, but they are able to move their bodies in complex ways. Their body language can also be read by human beings,

indicating arousal states, attentional directions, and overall wellness. Though dogs cannot point like children can, they may be using particular body movements or orientations prior to instances of joint attention. Careful analysis of body movement and orientation can reveal more about joint attention in this cross-species context and this can provide further comparison to joint attention in children.

References

Abney, D. H., Smith, L. B., & Yu, C. (2017). It's Time: Quantifying the Relevant Timescales for Joint Attention. *Proceedings of the Annual Meeting of the Cognitive Science Society*. https://doi.org/https://escholarship.org/uc/item/3rk9n2v8

Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants Parse Dynamic Action. *Child Development*, *72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

Bradley, H., Smith, B. A., & Wilson, R. B. (2023). Qualitative and Quantitative Measures of Joint Attention Development in the First Year of Life: A Scoping Review. *Infant and Child Development*, *32*(4), e2422. https://doi.org/10.1002/icd.2422

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. https://doi.org/10.1007/s10071-007-0088-1

Byrne, M., Horschler, D. J., Schmitt, M., & Johnston, A. M. (2023). Pet dogs (*Canis familiaris*) re-engage humans after joint activity. *Animal Cognition*, *26*(4), 1277–1282. https://doi.org/10.1007/s10071-023-01774-1

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint Attention and Imitative Learning in Children, Chimpanzees and Enculturated Chimpanzees. *Social Development*, *4*(3), 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00510

Dale, R., Marshall-Pescini, S., & Range, F. (2020). What matters for cooperation? The importance of social relationship over cognition. *Scientific Reports*, *10*, 11778. https://doi.org/10.1038/s41598-020-68734-4

Dwyer, F., Bennett, P. C., & Coleman, G. J. (2006). Development of the Monash Dog Owner Relationship Scale (MDORS). *Anthrozoös*, *19*(3), 243–256. https://doi.org/10.2752/089279306785415592

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Developmental science*, *21*(4), e12626. https://doi.org/10.1111/desc.12626

Gabouer, A., & Bortfeld, H. (2021). Revisiting how we operationalize joint attention. *Infant behavior & development*, *63*, 101566. https://doi.org/10.1016/j.infbeh.2021.101566

Gao, H., Wang, W., Cai, M., Wang, D., Gao, W., & Cai, J. (2025). Unveiling the Nexus: How Perceived Intelligence Shapes the Willingness to Use Robot. *International Journal of Human–Computer Interaction*, *41*(24), 15915–15924. https://doi.org/10.1080/10447318.2025.2504181

Gaunet, F., & Deputte, B. L. (2011). Functionally referential and intentional communication in the domestic dog: Effects of spatial and social contexts. *Animal Cognition*, *14*(6), 849–860. https://doi.org/10.1007/s10071-011-0418-1

Guo, J., & Feng, G. (2013). How Eye Gaze Feedback Changes Parent-Child Joint Attention in Shared Storybook Reading? In Y. I. Nakano, C. Conati, & T. Bader (Eds.). Springer London. https://doi.org/10.1007/978-1-4471-4784-8_2

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151. https://doi.org/10.1006/anbe.2000.1518

Heesen, R., Bangerter, A., Zuberbühler, K., Rossano, F., Iglesias, K., Guéry, J.-P., & Genty, E. (2020). Bonobos engage in joint commitment. *Science Advances*, *6*(51), eabd1306. https://doi.org/10.1126/sciadv.abd1306

Hermes, J., Behne, T., Studte, K., Zeyen, A.-M., Gräfenhain, M., & Rakoczy, H. (2016). Selective Cooperation in Early Childhood – How to Choose Models and Partners. *PLOS ONE*, *11*(8), e0160881. https://doi.org/10.1371/journal.pone.0160881

Hopper, L. M., & Cronin, K. A. (2018). What Did You Get? What Social Learning, Collaboration, Prosocial Behaviour, and Inequity Aversion Tell Us About Primate Social Cognition. In L. D. Di Paolo, F. Di Vincenzo, & F. De Petrillo (Eds.), *Evolution of Primate Social Cognition* (pp. 13–26). Springer International Publishing. https://doi.org/10.1007/978-3-319-93776-2_2

Horschler, D. J., Bray, E. E., Gnanadesikan, G. E., Byrne, M., Levy, K. M., Kennedy, B. S., & MacLean, E. L. (2022). Dogs re-engage human partners when joint social play is interrupted: A behavioural signature of shared intentionality? *Animal Behaviour*, *183*, 159–168. https://doi.org/10.1016/j.anbehav.2021.11.007

Howell, T. J., Bowen, J., Fatjó, J., Calvo, P., Holloway, A., & Bennett, P. C. (2017). Development of the cat-owner relationship scale (CORS). *Behavioural Processes*, *141*(Pt 3), 305–315. https://doi.org/10.1016/j.beproc.2017.02.024

Huang, X., & Lajoie, S. P. (2023). Social emotional interaction in collaborative learning: Why it matters and how can we measure it? *Social Sciences & Humanities Open*, *7*(1), 100447. https://doi.org/10.1016/j.ssaho.2023.100447

Jermann, P., Nüssli, M.-A., & Li, W. (2010). Using dual eye-tracking to unveil coordination and expertise in collaborative Tetris. *Proceedings of HCI 2010*. https://doi.org/10.14236/ewic/HCI2010.7

Jonkoski, D. S., Douglas, L. E. L. C., Horschler, D. J., Klensin, A. M., Kennedy, B. S., MacLean, E. L., & Bray, E. E. (2026). Indiscriminate sociality: Puppies do not preferentially re-engage a human partner after joint social play is interrupted. *Animal Behaviour*, *231*, 123425. https://doi.org/10.1016/j.anbehav.2025.123425

Kaplan, F., & Hafner, V. V. (2006). The challenges of joint attention. *Interaction Studies*, *7*(2), 135–169. https://doi.org/10.1075/is.7.2.04kap

Liao, S., Lin, L., Chen, Q., & Pei, H. (2024). Why not work with anthropomorphic collaborative robots? The mediation effect of perceived intelligence and the moderation effect of self-efficacy [_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hfm.21024]. *Human Factors and Ergonomics in Manufacturing & Service Industries*, *34*(3), 241–260. https://doi.org/10.1002/hfm.21024

Malle, B. F. (2008). The Fundamental Tools, and Possibly Universals, of Human Social Cognition. In *Handbook of Motivation and Cognition Across Cultures* (1st ed., pp. 267–296). Elsevier.

Marshall-Pescini, S., Colombo, E., Passalacqua, C., Merola, I., & Prato-Previde, E. (2013). Gaze alternation in dogs and toddlers in an unsolvable task: Evidence of an audience effect. *Animal Cognition*, *16*(6), 933–943. https://doi.org/10.1007/s10071-013-0627-x

Merola, I., Prato-Previde, E., & Marshall-Pescini, S. (2012). Social referencing in dog-owner dyads? *Animal Cognition*, *15*(2), 175–185. https://doi.org/10.1007/s10071-011-0443-0

Monroy, C., Chen, C.-H., Houston, D., & Yu, C. (2020). Action prediction during real-time parent-infant interactions. *Developmental Science*, *24*, e13042. https://doi.org/https://doi.org/10.1111/desc.13042

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022). Head-mounted mobile eye-tracking in the domestic dog: A new method. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01907-3

Pelgrim, M. H., Raman, S. S., Serre, T., & Buchsbaum, D. (2025). Evaluating Dogs' Real-World Visual Environment and Attention. *Cognitive Science*, *49*, e70080. https://doi.org/10.1111/cogs.70080

Pelgrim, M. H., Ross, M. A., Malle, B. F., & Buchsbaum, D. (2026). *Human Judgements of Dog Intelligence* (Manuscript in preparation).

Pelgrim, M. H., Tidd, Z., Byrne, M., Johnston, A. M., & Buchsbaum, D. (2024). Synchronous citizen science with dogs. *Animal Cognition*, *27*(1), 46. https://doi.org/10.1007/s10071-024-01882-6

Povinelli, D. J., & Eddy, T. J. (1996). What Young Chimpanzees Know about Seeing. *Monographs of the Society for Research in Child Development*, *61*(3), i–189. https://doi.org/10.2307/1166159

Range, F., Kassis, A., Taborsky, M., Boada, M., & Marshall-Pescini, S. (2019). Wolves and dogs recruit human partners in the cooperative string-pulling task. *Scientific Reports*, *9*(1), 17591. https://doi.org/10.1038/s41598-019-53632-1

Ross, H. S., & Lollis, S. P. (1987). Communication within infant social games. *Developmental Psychology*, *23*(2), 241–248. https://doi.org/10.1037/0012-1649.23.2.241

Ross, M., Buchsbaum, D., & Malle, B. F. (2024). Human Perceptions of Canine Intelligence. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*. https://escholarship.org/uc/item/3d9549vn

Shvarts, A. (2018). Joint Attention in Resolving the Ambiguity of Different Presentations: A Dual Eye-Tracking Study of the Teaching-Learning Process. In N. Presmeg, L. Radford, W.-M. Roth, & G. Kadunz (Eds.), *Signs of Signification: Semiotics in Mathematics Education Research* (pp. 73–102). Springer International Publishing. https://doi.org/10.1007/978-3-319-70287-2_5

Smith, B. P., & Litchfield, C. A. (2013). Looking back at 'looking back': Operationalising referential gaze for dingoes in an unsolvable task. *Animal Cognition*, *16*(6), 961–971. https://doi.org/10.1007/s10071-013-0629-8

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Wallot, S., & Leonardi, G. (2018). Analyzing Multivariate Dynamics Using Cross-Recurrence Quantification Analysis (CRQA), Diagonal-Cross-Recurrence Profiles (DCRP), and Multidimensional Recurrence Quantification Analysis (MdRQA) – A Tutorial in R. *Frontiers in Psychology*, *9*, 2232. https://doi.org/10.3389/fpsyg.2018.02232

Warneken, F., Chen, F., & Tomasello, M. (2006). Cooperative Activities in Young Children and Chimpanzees. *Child Development*, *77*(3), 640–663. https://doi.org/10.1111/j.1467-8624.2006.00895.x

Warneken, F., Gräfenhain, M., & Tomasello, M. (2012). Collaborative partner or social tool? New evidence for young children's understanding of joint intentions in collaborative activities. *Developmental Science*, *15*(1), 54–61. https://doi.org/10.1111/j.1467-7687.2011.01107.x

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659

CHAPTER 6

Conclusion

6.1 Summary of Findings

In this thesis, I have discussed a collection of work examining how humans and dogs visually interact with each other.

Chapter 2 established a baseline understanding of how dogs look at the world around them on an ecologically valid task, specifically a dog walk. This study used wearable eye-tracking in a novel outdoor context. We integrated computer vision techniques to capture the item categories available for dogs to look at, as well as how they looked at the items. Chapter 2 showed how dogs preferentially direct their visual attention to specific item categories such as people and construction equipment. We found that dogs direct their visual attention to traversable ground in front of them at similar rates to other species. Unexpectedly, dogs were highly visually engaged in plant material like trees and bushes in their environment. On screen-based eye-tracking tasks these item categories are frequently used as background, and it is interesting that in the real world they appear to be visually interesting in their own right. These findings may be influenced by the potential integration of visual and olfactory information in dogs' minds (Andrews et al., 2022).

We integrated visual saliency analyses to show that dogs' visual behaviors are poorly explained by a strictly bottom-up low-level visual saliency model. This Chapter provided a baseline understanding of how dogs look at their world in a context where they are accompanied by a human, but not necessarily directed by them.

In Chapter 3, we examined how toddlers, dogs, and adults understood pointing gestures produced by an experimenter on a 4-alternative object search task. This chapter expanded our understanding of point-following beyond the common 2-alternative tasks used in prior work. We integrated depth tracking cameras to examine how both species approached the problem, supplementing our typical choice data. Dogs followed points at levels significantly different from chance but far below ceiling. Dogs' choices were biased towards the inner locations, choosing the locations closer to the pointer more often than the locations on either far side. Toddlers were better at following the point than dogs, outperforming them at finding the hidden item. Toddlers also chose between the four potential locations at approximately equal rates. However, toddlers made numerous errors. When examining choice behavior by side of the experimenter rather than specific item (which adjusted for dogs' bias for the inner cups), the difference between the two species choices disappeared. Together this suggests that toddlers do not have similar biases on this task as dogs.

In addition to the differences in choices between dogs and toddlers, we also observed species differences in path approach dynamics. In keeping with their preference for the inner locations, dogs followed more closely to the optimal path when first selecting one of the two cups closest to the experimenter. In contrast, toddlers followed more closely to the optimal path when first selecting one of the outer locations, specifically the outer location on the experimenter's left side. Future work might explore why this pattern was observed.

Study 2 of Chapter 3 demonstrated that adults were nearly perfect at following the point in an adapted online version of the task. This provides confirmation that the task itself is solvable, and provides a developmental end-point for comparison for our toddlers. Chapter 3 suggests that both toddlers and dogs are capable of following points produced by a relatively unknown

experimenter, but this is not a trivial task. At least for the toddlers, there is still significant development that happens to achieve adult performance levels.

Turning to the human components of the previous task, Chapter 4 continued to examine how people and their dogs interact in a more naturalistic way. Chapter 4 aimed to balance the naturalistic properties of Chapter 2 with the ability to gather controlled data seen in Chapter 3. On a 4-alternative object search task, we captured how dog guardians naturally pointed for their dogs. Using insights from robotics and computer science, this study examined potential candidate vectors for quantifying and representing pointing gestures produced by people. We found that the vector from the eye to the wrist of the freely moving dog owners was the most successful vector, as measured by three different accuracy metrics. Put another way, the gestures people naturally produce for their dogs integrate gaze and arm movement to direct their dogs to targets. Our results suggest that gaze and "pointing" at least for distant targets are potentially used as a single integrated cue rather than as two distinct additive ones. This may help to explain why prior point following tasks with dogs where gaze cues conflict or are interrupted have found such poor results (ManyDogs et al., 2023).

The findings of Chapter 4 suggest that dog owners may view pointing for their dogs as a bid or an initiation for joint attention. Specifically, owners were orienting their visual attention and gaze in line with their hand gesture to frame the referent of their point.

Quantitatively, we saw differences in our candidate vector measures of Euclidean distance and weighted accuracy between owners (Chapter 4) and experimenters (Chapter 3). Further, of our candidate vectors produced by experimenters, the Shoulder-Wrist vector was our best performing as measured by two of our accuracy measures. Our Elbow-Wrist was our best performer by our third accuracy measure. This is a completely different pattern than observed in naturalistic pointing produced by dog owners, where the Eye-Wrist vector was consistently the best.

Although differences in pointing gestures were observed, in our context they did not appear

to hugely bias choice performance. Dogs chose similarly in Chapter 3 and Chapter 4. As with points produced by experimenters, dogs can follow points produced by their owners to find a hidden treat at levels significantly above chance, they struggle with the task and are far from ceiling. This is likely because of their demonstrated preference for the inner cups, meaning those closer to their owner. This is the same pattern observed for dogs following the point from an experimenter outlined in Chapter 3, and dogs choice patterns are remarkably similar across the two studies. Dogs first choices were correct 39.7% of the time when following an experimenter (Chapter 3) and 41% of the time when following their owner (Chapter 4). Interestingly, across Chapters 3 and 4 we found that dogs take closer to optimal paths towards their first choices when following points produced by their owners than when following points produced by the experimenters. We also saw more optimal paths when selecting the inner locations, in this case the cups by the owner, than the outer ones in both Chapters. Further, we saw less variance in path deviation when following a point produced by an owner.

In Chapter 5, we aimed to maximize (within a lab setting) ecological validity, expanding on the dynamics explored in Chapter 4. People and their dogs worked together to locate hidden treats in a relatively unconstrained object-search task in a highly complex environment. We alternated which member of the pair had knowledge about the treats location. We used a dual eye-tracking approach similar to that previously used with human children and their parents to examine how people and their dogs visually coordinated their attention during the task. Chapter 5 explicitly tested theories of human-unique visual interactions, finding instances of both joint attention and mutual gaze occurring during search behaviors.

We found that dogs and people visually coordinated their attention to a greater degree than would be expected from a generated random baseline. Further, we saw instances of joint attention and mutual gaze across all of our dyads. Dyads were generally successful at our search task, locating the treat within the trial time limit successfully the majority of the time. Key visual behaviors, specifically joint attention, mutual gaze, and the interaction of the two predicted success on our search task. Joint attention in dyads, serving to promote task success, is a

direct challenge of the shared intentionality hypothesis. We also found instances of other gaze behaviors previously observed from behavioral coding. Gaze alternation and referential gazing occurred in all of our dyads.

Chapter 5 also highlighted that joint attention in dyads looks qualitatively different than in parent-child interactions (Yu & Smith, 2013). We examined the lag time between when the initiator of the joint attention began looking to the object and when their partner began to look to the object, thus starting joint attention. Our distribution of lag times was not symmetrical, and joint attention bouts were generally initiated by dogs.

6.2 Limitations

Newer technologies like head-mounted eye-tracking from Chapters 2 and 5 allow researchers to quantify visual attention and gain a rich data set during relatively natural events. However, a challenge of this particular technology is the generalizability of the findings. More specifically, the kinds of dyads willing to train to wear goggles are not necessarily representative of the general pet owner population. Training dogs to wear the goggles typically takes around a month (Pelgrim et al., 2022), and the goggles themselves are size and shape limited. It is possible that brachycephalic dogs (flat-faced dogs without extended snouts who cannot wear goggles) may see the world differently. This will remain an open question until hardware is updated to be compatible with such skull morphologies.

Individual differences between dogs and human-dog dyads likely also play a large role in their visual interaction patterns. A consistent challenge for eye-tracking studies like those mentioned here is gathering sufficient sample sizes. Although our studies are adequately powered for the specific claims we make, this comes with the continued limitation that training backgrounds and prior experiences undoubtedly shape the way dogs look at the world, something not explored here. For example, an owner of a pet dog who knows many tricks, and the owner of a dog working in search and rescue may both be willing to participate in goggle training. Further,

both dogs may be tolerant of goggles and sufficiently food motivated to participate in studies. However, these two dyads are likely different in their interactions with their environments and with each other. Whether resulting from prior reward history, dog temperament, or a combination of the two, individual differences certainly exist between the way people and dogs approach the world.

People who are willing to participate in studies with their dogs are generally highly bonded to their dogs and feel emotionally close to them. Although owners in our sample are consistent with general dog owners in feeling their dogs are above average in intelligence, we must be careful not to draw sweeping conclusions about the general public from these findings. Dog owners participating in, for example, Chapter 5, had to be highly motivated and committed to their participation and to the research. Participants had to train their dogs to wear goggles, and then participated themselves in multiple surveys, and finally completed our task itself.

We have presented findings showing a species capability. The specific visual patterns and the role of joint attention and mutual gaze promoting task success may not be true in every possible human-dog interaction context. Not every dog may be as successful at working with a person in our task. In other collaborative tasks, dogs generally work well with people, but future work should explore if the visual behaviors we have seen (and the role they appear to play) extend to human-dog interaction more generally, independent of the relationship between that dyad (Range et al., 2019).

There is a large interest in potential breed or breed group differences between dogs. Most of the dogs in these studies are of mixed breed origin, so no claims can be made about differences in breed or breed group. This limitation is not unique to this thesis. Even large scale studies using big team science approaches have a difficult time exploring cognitive and behavioral differences between breed groups (ManyDogs et al., 2023). Future researchers seeking to better understand the differences between dog breeds or breed groups will need to seek out these populations and will likely be limited to make comparisons between a small number of breeds.

6.3 Implications and Future Directions

Toddlers' performance on our task in Chapter 3 suggests that point following at 17-24 months may not facilitated by as rich of a mechanism as predicted by prior theories. Pointing is often held as a rich and highly social communicative gesture that facilitates joint attention and other complex cognitive collaborative processes (Tomasello et al., 2007). From prior work, toddlers can be expected to readily engage in joint attention well before 17 months of age (Carpenter et al., 1995; Tomasello et al., 2005). Under a rich interpretation of point following, we should expect toddlers to perform near ceiling. The data presented here do not undermine prior evidence demonstrating joint attention in toddlers, rather they suggest that toddlers may default to approaching pointing gestures with a different cognitive mechanism.

Prior to joint attention, 4-month-old infants are capable of following object oriented points, meaning they preferentially look at an indicated referent (Bertenthal et al., 2014). It is possible that when the stakes are relatively low, like in this exhaustive search task, toddlers rely on the same lower-level mechanism they used earlier in life. Again, this does not suggest that the developed capacity for joint attention is somehow absent in this age group. Rather, potentially due to the cognitive effort required, they choose to use simpler and "lower-level" cues such as learned associations. Alternatively, toddlers may be engaging in joint attention and understand the task as directing shared intentionality, but they may have difficulty with inhibitory control that prevents them from making correct choices and taking optimal paths.

Future work can further investigate these potential explanations for toddlers challenges with this success in a few ways. First, future work could consider eliminating the exhaustive search approach. Toddlers were highly motivated to locate the hidden item, but they may have also found the searching activity rewarding. Eliminating the options for repeated searches could result in higher exclusion rates, as toddlers may become de-motivated in the event of repeated failures. However, by removing repeated search options the only rewarding item or goal is the item itself. This could help to resolve potential ambiguity about the goal of the task that may have also resulted in toddlers defaulting to lower-level point following cues. If toddlers interpret

pointing as a rich cue, we could expect that they would search and approach optimally following the joint intentionality initiated by the experimenter.

Potential difficulties with inhibitory control could also be explored by testing toddlers on an existing measure of inhibitory control (Carlson et al., 2004; Holmboe et al., 2018; Kochanska et al., 2000). If toddlers have a rich interpretation of pointing gestures and are having difficulty due to inhibitory control, we should see a strong correlation between better performance on the inhibitory control measures and choice success. In contrast, if there is no relationship between other validated measures of inhibitory control and task success it would further support the alternative that toddlers default to a lean interpretation of pointing gestures.

The difference in choice and approach data we saw in toddlers and dogs suggests that the two species approach the problem differently, weighing their decisions on the basis of different factors. There also may be a different cognitive mechanism underlying the point following behavior seen in both groups. Dogs are undoubtedly biased by the proximity of the experimenter, but barring that bias both species appear to make errors at similar rates. Both dogs and toddlers may be approaching this task with a relatively low-level interpretation of the meaning of the experimenters' pointing gesture.

In ongoing work using eye-tracking we are exploring what both species are looking at when solving this problem. This can help us to better understand how the various gestures produced by pointers are processed and weighed by the viewer over the course of a trial.

People and dogs work together in a variety of capacities. We frequently interact with dogs, and, as visual creatures, we come primed to interact with others through visual means (Ellsworth & Ludwig, 1972; Kaas & Balaram, 2014). In keeping with prior work (Bensky et al., 2013), although dogs' primary sensory modality may not be vision, Chapter 2 demonstrates that dogs do actively engage with their world visually. Their visual attention is directed selectively and cannot be explained by low-level saliency models. Intuitively, this makes sense in light of prior findings of point comprehension. Following points, as demonstrated in prior work and in Chapter 3 and 4,

would not be possible without dogs visually assessing their social partners' cues. When engaged in a shared or combined task, dogs and humans visually coordinate their attention. Demonstrated in Chapter 5, dogs both lead and follow the visual attention of their human partners. Humans and dogs actively engage with each other in a visual manner, coordinating their looking behavior to accomplish shared goals.

Future work must continue to expand on how people and their dogs interact and should continue to make advances in increasing ecological validity. Research should aim to explore a larger variety of dogs, capturing dogs writ large not just those who are trainable with highly motivated owners. Notably, experimenters in Chapter 3 appear to point differently than dog owners (Chapter 4). The subtle differences we observed in dog performance, in combination with the large differences observed in pointer performance, suggest a need for future, careful, consideration when attempting to draw conclusions about how dogs behave in the real-world from tightly controlled experimental settings.

Researchers in the future should further consider expanding the experiments and findings presented here into new populations, particularly dogs in working roles. This dissertation tested pet dogs of a variety of training backgrounds. Future work should consider testing dogs in working roles or those in training. Current training programs for dogs in working and service roles are expensive and have high failure rates. Identifying what makes dogs or makes teams successful at the task at hand can improve earlier selection and training procedures more broadly. For example, future work might consider using the dual eye-tracking paradigm from Chapter 5 to explore how search and rescue dogs work with their handlers. Identifying key aspects of team dynamics that promote task success could be used to improve handler and dog training, making teams more efficient at their life-saving work.

There is also a large interest in better understanding human-dog interactions for promoting successful human-robot interactions (Byrne et al., 2020; Krueger et al., 2021; Szabó et al., 2010). In ongoing work, we are examining how the data collected on human-dog interactions can be used to improve robot search algorithms. Training robots for human interactions, with

dogs in mind, shows particular promise, in part because people are generally successful at interacting with dogs. In part because of dogs' ubiquity in our environment, many robots are built to be dog-like in appearance (Faragó et al., 2014; Kerepesi et al., 2006). In efforts to extend the findings presented here, future research could consider further exploring how dogs and relative strangers work together. Designers of robots aim for their products or algorithms to be intuitive and usable by the public. In our relatively controlled tasks (like in Chapter 3 and 4), we have demonstrated that dogs can follow points from a relative stranger as well as their owner. However, we also observed subtle behavioral differences in the way that dogs approached the task.

Unlike in pointing tasks, it remains unknown how dogs would approach the complex collaborative search task presented in Chapter 5 with someone other than their owner. It is possible that dogs would be equally successful with a stranger, and that they would display consistent visual behaviors. Alternatively, it is possible that dogs and their owners (or at least a person the dog is familiar with) have familiar gestures or communication styles that allow them to be more successful on the task. Existing dyads may be capitalizing on behavioral shorthands, facilitating success on tasks with reduced effort compared to a novel pair. As an anecdote, in Chapter 5, one dog owner noted that she made an error during the first collaborative trial by saying "find it", which she generally uses to mean search and find multiple items. Behaviorally, this was reflected in the dog continuing to search after the trial had ended. During subsequent trials, the owner used the terms "show me" or "go get it", and we did not observe continued post-trial searching. This anecdote highlights that dogs and their owners regularly engage in multi-modal (audio-visual) communication. It also highlights how dogs' prior experiences with their owners can facilitate collaborative searching. Future research can consider analyzing this further, exploring how dogs integrate auditory cues with visual ones. Additionally, future work can explore whether the behavioral shorthands typically used by dogs and their owners are multi-modal (including words and gestures) or unimodal (such as a hand sign or single vocalization).

Multi-modal perception also occurs using senses we are less reliant on, like smell. Dogs

are often thought of as olfactory creatures. We know from numerous studies that dogs have an exceptional sense of smell. They are able to detect and differentiate between small concentrations of odors and can use these odors to generalize and create categories or concepts (Hall et al., 2016; Lazarowski et al., 2019). When interacting independently with the world or when trained to do so, dogs may prioritize smell cues. However, when working with a human partner, they rely on their vision (Bensky et al., 2013). This is even true to their own detriment. Dogs follow human social-communicative cues perceived visually, and consistently fail to use olfactory cues to avoid misleading human information (Szetei et al., 2003). Existing literature has focused on olfaction in working dog populations or from a neural perspective (Andrews et al., 2022; Guran et al., 2024; Lazarowski et al., 2020). The collection of studies presented here have focused on dog visual attention, but future research can explore how olfaction may also play a role in human-dog interactions.

The findings of this thesis challenge established theories about human-unique cognitive abilities, specifically the shared intentionality hypothesis. The shared intentionality hypothesis argues that human social learning and interaction is facilitated by a specific mechanism - an ability to form goals and intentions, and share them with others (Tomasello et al., 2005). Further, it claims that the observable behavioral signature of this shared intentionality is in joint visual attention towards a shared item.

The focal comparison species used to identify differences between humans and non-human animals was chimpanzees (Gardner, 2005; Tomasello et al., 2005). Testing for rich social abilities in chimpanzees may have resulted in overestimations of human unique traits. Chimpanzees are not the most cooperatively social of the great apes, tending to perform best in competitive contexts (Bräuer et al., 2007; Hare et al., 2001). In contrast, focusing on highly cooperative species has shown that we are not the only ones to engage in the key observable behaviors of triadic engagement. Bonobos (another great ape species that is highly cooperative and social) readily re-engage social partners when an activity was interrupted (Heesen et al., 2020). Bonobos also cooperate with conspecifics outside of their social group or community (Samuni & Surbeck,

2023).

Even among our own species, the shared intentionality hypothesis has remained controversial. The developmental timeline requires a lean interpretation of infant looking time work on goals and intentions. If looking time tasks are taken at face value, they suggest that infants understand goal sharing earlier than the the 9-12 month timeline expected with the emergence of triadic engagement (Vaish & Woodward, 2005). Further, recent neural evidence suggests infants may understand and recognize collaborative engagement in others by 9 months of age, earlier than the expected 12-15 month window (Begus et al., 2020).

Supporters of the shared intentionality hypothesis have failed to seriously address field and observational data on their chosen comparison species (Boesch, 2005). Chimpanzees, like many other species, hunt collaboratively in groups. The focus on operationalizing joint intentions and shared goals as specific observable visual behaviors may have also contributed to an underestimation of the abilities of other species. There may also have been an overestimation of the extent of joint attention in young children, or at least the functional benefits.

The shared intentionality hypothesis would suggest that well before 17 months (the age of our youngest human participants in Chapter 3) toddlers should have a rich and highly social understand pointing gestures and be highly successful at following them. The pointing gesture, in view of this theory, is used in conjunction with other visual cues to work together with joint intentions to engage in a collaborative task. This theory was weakened by the challenges observed by the toddlers in Chapter 3. As discussed earlier, these findings do not undo prior evidence that toddlers are capable of creating joint intentions and shared goals with social partners. They suggest that it may not be the default way they approach tasks like point following.

Toddlers in our sample were given full ostensive cues (name calling, initial eye-contact by the experimenter) and the pointing gesture was present for the entire search process (reducing working memory load as the social information was always available). Toddlers may be falling

back on a lower-level cognitive mechanism to interpret point following. They may also choose to selectively engage in collaborative engagement with the pointer, perhaps because it is challenging or requires more inhibitory control. This could explain how they are sometimes successful and why their errors are relatively random. Future work is needed to better understand why children are challenged by a 4-alternative task, and the level at which they naturally use collaborative engagement with social partners on day to day tasks. Ongoing work using head-mounted eye-tracking will reveal more information about how the pointing gestures are being processed by the two species. However, at present toddlers relatively poor performance compared to adults and the ability for dogs to respond to the points in this context pose significant challenges for the shared intentionality hypothesis.

Findings from Chapter 3 and 4 suggest that dogs are capable of following pointing gestures at above chance levels to distant referents, something that should be only discernible through triadic engagement (Lyn & Christopher, 2020). Consistent patterns of visual re-engagement from adult dogs (Byrne et al., 2023; Horschler et al., 2022) and from bonobos (Heesen et al., 2020) further support the idea that the operationalizations of triadic engagement are not unique to our species.

Dogs' ability to cooperate with human partners in a flexible and selective way, as well as their ability to engage in joint attention, as seen in Chapter 5, should be unique to human beings. Under the shared intentionality hypothesis, joint attention involves coordinating their visual focus towards the same real-world entity and is taken to be representative of shared goals resulting in their joint intentions to execute an action plan (Tomasello et al., 2005). Observing this specific visual behavior in our dog-human pairs has demonstrated that this visual pattern of behavior is not unique to our species.

It is perhaps unnecessary to state explicitly, but human beings are able to collaborate with each other in unique ways. No other species can work together to build and create the way we can, but the findings of this thesis suggest that we have set the bar too low when seeking the cognitive abilities that make us unique. Given that another species has now been demonstrated to

engage in joint attention as it has been previously operationalized in children, we must consider where to go next.

Shared intentionality may truly be necessary and sufficient to explain human unique cognition. If so, we must reconsider the visual behavioral correlates used to identify the stages. Across multiple operationalizations, joint attention is seen and initiated by a non-human animal. Perhaps the full suite of cognitive abilities behind joint intentionality are unique to humans, but the visual behavior we have previously accepted as joint attention is shared. It is also possible that joint attention could exist on a spectrum. Maybe joint attention in our human-human interactions is different from what we see in this cross-species context. There may be subtle differences in the patterns of our joint attention that we could use to distinguish from those seen in dogs. This would require careful operationalization of how we are defining gradients of joint attention, avoiding subjective judgments and attributions of intentionality. Finally, shared intentionality may just be one of a suite of abilities that can explain our unique cognition. This thesis suggests that in defining what makes us unique we must re-operationalize or revise the shared intentionality hypothesis.

References

Andrews, E. F., Pascalau, R., Horowitz, A., Lawrence, G. M., & Johnson, P. J. (2022). Extensive Connections of the Canine Olfactory Pathway Revealed by Tractography and Dissection. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *42*(33), 6392–6407. https://doi.org/10.1523/JNEUROSCI.2355-21.2022

Begus, K., Curioni, A., Knoblich, G., & Gergely, G. (2020). Infants understand collaboration: Neural evidence for 9-month-olds' attribution of shared goals to coordinated joint actions. *Social Neuroscience*, *15*(6), 655–667. https://doi.org/10.1080/17470919.2020.1847730

Bensky, M. K., Gosling, S. D., & Sinn, D. L. (2013, January). Chapter Five - The World from a Dog's Point of View: A Review and Synthesis of Dog Cognition Research. In H. J. Brockmann, T. J. Roper, M. Naguib, J. C. Mitani, L. W. Simmons, & L. Barrett

(Eds.), *Advances in the Study of Behavior* (pp. 209–406, Vol. 45). Academic Press. https://doi.org/10.1016/B978-0-12-407186-5.00005-7

Bertenthal, B. I., Boyer, T. W., & Harding, S. (2014). When do infants begin to follow a point? *Developmental Psychology*, *50*(8), 2036–2048. https://doi.org/10.1037/a0037152

Boesch, C. (2005). Joint cooperative hunting among wild chimpanzees: Taking natural observations seriously. *Behavioral and Brain Sciences*, *28*(5), 692–693. https://doi.org/10.1017/S0140525X05230121

Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition*, *10*(4), 439–448. https://doi.org/10.1007/s10071-007-0088-1

Byrne, C., Logas, J., Freil, L., Allen, C., Baltrusaitis, M., Nguyen, V., Saad, C., & Jackson, M. M. (2020). Dog Driven Robot: Towards Quantifying Problem-Solving Abilities in Dogs. *Proceedings of the Sixth International Conference on Animal-Computer Interaction*, 1–5. https://doi.org/10.1145/3371049.3371063

Byrne, M., Horschler, D. J., Schmitt, M., & Johnston, A. M. (2023). Pet dogs (*Canis familiaris*) re-engage humans after joint activity. *Animal Cognition*, *26*(4), 1277–1282. https://doi.org/10.1007/s10071-023-01774-1

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive Function and Theory of Mind: Stability and Prediction From Ages 2 to 3. *Developmental Psychology*, *40*(6), 1105–1122. https://doi.org/10.1037/0012-1649.40.6.1105

Carpenter, M., Tomasello, M., & Savage-Rumbaugh, S. (1995). Joint Attention and Imitative Learning in Children, Chimpanzees and Enculturated Chimpanzees. *Social Development*, *4*(3), 217–237. https://doi.org/10.1111/j.1467-9507.1995.tb00063.x

Ellsworth, P. C., & Ludwig, L. M. (1972). Visual Behavior in Social Interaction. *Journal of Communication*, *22*(4), 375–403. https://doi.org/10.1111/j.1460-2466.1972.tb00164.x

Faragó, T., Gácsi, M., Korcsok, B., & Miklósi, Á. (2014). Why is a dog-behaviour-inspired social robot not a doggy-robot? *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, *15*(2), 224–232. https://doi.org/10.1075/is.15.2.11far

Gardner, R. A. (2005). Animal Cognition Meets Evo-Devo. *Behavioral and Brain Sciences*, *28*(5), 699–700. https://doi.org/10.1017/s0140525x05310120

Guran, C.-N. A., Boch, M., Sladky, R., Lonardo, L., Karl, S., Huber, L., & Lamm, C. (2024). Functional mapping of the somatosensory cortex using noninvasive fMRI and touch in awake dogs. *Brain Structure and Function*, *229*(5), 1193–1207. https://doi.org/10.1007/s00429-024-02798-0

Hall, N. J., Collada, A., Smith, D. W., & Wynne, C. D. L. (2016). Performance of domestic dogs on an olfactory discrimination of a homologous series of alcohols. *Applied Animal Behaviour Science*, *178*, 1–6. https://doi.org/10.1016/j.applanim.2016.03.016

Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, *61*(1), 139–151. https://doi.org/10.1006/anbe.2000.1518

Heesen, R., Bangerter, A., Zuberbühler, K., Rossano, F., Iglesias, K., Guéry, J.-P., & Genty, E. (2020). Bonobos engage in joint commitment. *Science Advances*, *6*(51), eabd1306. https://doi.org/10.1126/sciadv.abd1306

Holmboe, K., Bonneville-Roussy, A., Csibra, G., & Johnson, M. H. (2018). Longitudinal development of attention and inhibitory control during the first year of life. *Developmental Science*, *21*(6), e12690. https://doi.org/10.1111/desc.12690

Horschler, D. J., Bray, E. E., Gnanadesikan, G. E., Byrne, M., Levy, K. M., Kennedy, B. S., & MacLean, E. L. (2022). Dogs re-engage human partners when joint social play is interrupted: A behavioural signature of shared intentionality? *Animal Behaviour*, *183*, 159–168. https://doi.org/10.1016/j.anbehav.2021.11.007

Kaas, J. H., & Balaram, P. (2014). Current research on the organization and function of the visual system in primates. *Eye and Brain*, *6*(sup1), 1–4. https://doi.org/10.2147/EB.S64016

Kerepesi, A., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., & Miklósi, Á. (2006). Behavioural comparison of human–animal (dog) and human–robot (AIBO) interactions. *Behavioural Processes*, *73*(1), 92–99. https://doi.org/10.1016/j.beproc.2006.04.001

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology*, *36*(2), 220–232. https://doi.org/10.1037/0012-1649.36.2.220

Krueger, F., Mitchell, K. C., Deshpande, G., & Katz, J. S. (2021). Human–dog relationships as a working framework for exploring human–robot attachment: A multidisciplinary review. *Animal Cognition*. https://doi.org/10.1007/s10071-021-01472-w

Lazarowski, L., Krichbaum, S., DeGreeff, L. E., Simon, A., Singletary, M., Angle, C., & Waggoner, L. P. (2020). Methodological Considerations in Canine Olfactory Detection Research. *Frontiers in Veterinary Science*, *7*. https://doi.org/https://doi.org/10.3389/fvets.2020.00408

Lazarowski, L., Rogers, B., Waggoner, L. P., & Katz, J. S. (2019). When the nose knows: Ontogenetic changes in detection dogs' (*Canis familiaris*) responsiveness to social and olfactory cues. *Animal Behaviour*, *153*, 61–68. https://doi.org/10.1016/j.anbehav.2019.05.002

Lyn, H., & Christopher, J. (2020). A point is not a point is not a point: Reinterpreting three basic kinds of point comprehension. *The Evolution of Language: Proceedings of the 13th International Conference (Evolang13)*, 260–263. https://doi.org/10.17617/2.3190925

ManyDogs, Espinosa, J., Stevens, J. R., Alberghina, D., Barela, J., Bogese, M., Bray, E., Buchsbaum, D., Byosiere, S.-E., Cavalli, C., Dror, S., Fitzpatrick, H., Freeman, M. S., Frinton, S., Gnanadesikan, G. E., Guran, C.-N. A., Glover, M., Hare, B., Hare, E., … Walsh, C. (2023). ManyDogs 1: A multi-lab replication study of dogs' pointing comprehension. *Animal Behavior and Cognition*, *10*(3), 232–286. https://doi.org/https://doi.org/10.26451/abc.10.03.03.2023

Pelgrim, M. H., Espinosa, J., & Buchsbaum, D. (2022). Head-mounted mobile eye-tracking in the domestic dog: A new method. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01907-3

Range, F., Kassis, A., Taborsky, M., Boada, M., & Marshall-Pescini, S. (2019). Wolves and dogs recruit human partners in the cooperative string-pulling task. *Scientific Reports*, *9*(1), 17591. https://doi.org/10.1038/s41598-019-53632-1

Samuni, L., & Surbeck, M. (2023). Cooperation across social borders in bonobos. *Science*, *382*(6672), 805–809. https://doi.org/10.1126/science.adg0844

Szabó, C., Róka, A., Gácsi, M., Miklósi, Á., Baranyi, P., & Korondi, P. (2010). An emotional engine model inspired by human-dog interaction. *2010 IEEE International Conference on Robotics and Biomimetics*, 567–572. https://doi.org/10.1109/ROBIO.2010.5723388

Szetei, V., Miklósi, Á., Topál, J., & Csányi, V. (2003). When dogs seem to lose their nose: An investigation on the use of visual and olfactory cues in communicative context between dog and owner. *Applied Animal Behaviour Science*, *83*(2), 141–152. https://doi.org/10.1016/S0168-1591(03)00114-X

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A New Look at Infant Pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Vaish, A., & Woodward, A. (2005). Baby steps on the path to understanding intentions. *Behavioral and Brain Sciences*, *28*(5), 717–718. https://doi.org/10.1017/S0140525X05500128

Yu, C., & Smith, L. B. (2013). Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE*, *8*(11), e79659. https://doi.org/10.1371/journal.pone.0079659